

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:25:11 ; Search time 46 Seconds
(without alignments)
3036.508 Million cell updates/sec

Title: US-09-743-674-2
Perfect score: 4404
Sequence: 1 MPAGRLPRRCPMWTKFTDCT.....EEAEGDVLKRAQALQELGI 880

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4404	100.0	880	21	AAV44638 N. meningitidis T-
2	1122.5	25.5	244	24	ABP78566 N. gonorrhoeae ami
3	1122.5	25.5	244	24	ABP80077 N. gonorrhoeae ami
4	1025	23.3	275	24	ABP78509 N. gonorrhoeae ami
5	1025	23.3	275	24	ABP79684 N. gonorrhoeae ami
6	821.5	18.7	253	24	ABP78538 N. gonorrhoeae ami
7	620.5	14.1	132	24	ABP79950 N. gonorrhoeae ami
8	358	8.1	1095	24	ABU26247 Aspergillus fumiga
9	319.5	7.3	2768	22	ABB68397 Drosophila melanog

10	315.5	7.2	1786	18	AAW24790 P. falciparum live
11	315.5	7.2	1787	23	AAU96699 Plasmodium falcipo
12	315	7.2	1000	24	ABU25647 Aspergillus fumiga
13	313	7.1	1822	13	AAK27745 Extracellular fact
14	289.5	6.6	1468	22	ABB62991 Drosophila melanog
15	287.5	6.5	753	22	ABB68110 Drosophila melanog
16	277	6.3	2748	22	ABB58843 Drosophila melanog
17	276.5	6.3	2519	22	ABG16636 Novel human diagno
18	274.5	6.2	1192	22	ABBS9642 Drosophila melanog
19	267	6.1	688	21	AAV57274 E. canis 120 kDa i
20	256.5	5.8	8805	22	ABB67112 Drosophila melanog
21	253	5.7	710	22	ABU52992 Drosophila melanog
22	252	5.7	864	22	ABB71319 Drosophila melanog
23	252	5.7	6815	22	ABB66811 Drosophila melanog
24	247	5.6	842	22	ABB66631 Drosophila melanog
25	246	5.6	795	22	ABB68471 Plasmodium falcipa
26	243.5	5.5	1558	21	AAI18324 C. thermocellum Ol
27	239	5.4	1664	19	AAW43106 Drosophila melanog
28	238.5	5.4	654	22	ABB63266 Human liver peptid
29	238	5.4	2665	22	ABG48336 Human peptide #965
30	238	5.4	2665	22	ABB28314 Human peptide #965
31	238	5.4	2665	22	ABB33490 Peptide #996 encod
32	238	5.4	2665	22	ABBI8950 Protein #949 encod
33	238	5.4	2665	22	AAW54270 Human brain expres
34	238	5.4	2665	22	AAW66665 Human bone marrow
35	238	5.4	2665	22	AAW14533 Peptide #967 encod
36	238	5.4	2665	22	AAW26950 Peptide #987 encod
37	238	5.4	2665	22	AAW02259 Peptide #941 encod
38	238	5.4	2665	23	ABG36319 Human peptide enco
39	238	5.4	3266	21	AAW42491 Human ORF2255
40	238	5.4	3664	24	ABR47592 Breast cancer asso
41	237.5	5.4	1655	22	ABG61965 Drosophila melanog
42	236	5.4	2478	22	AAU34320 Staphylococcus aur
43	236	5.4	2478	22	AAU37374 Staphylococcus aur
44	236	5.4	2478	24	ABU19002 Pathogen specific
45	235	5.3	905	22	ABBS7827 Drosophila melanog

ALIGNMENTS

RESULT 1

AAV44638

ID AAY44638 standard; Protein; 880 AA.

XX

AC AAY44638;

XX

DT 18-APR-2000 (first entry)

XX

DE N. meningitidis T-cell stimulating protein A (Tspa).

XX

XX T-cell stimulating protein A; Tspa; CD4+ T-cell; stimulant; meningitis;
KW antibacterial; anti-inflammatory; vaccine; neisserial disease;
KW gonorrhoea; septicaemia; septic arthritis; pelvic inflammatory disease;
KW meningococcal; gonococcal.

XX

OS Neisseria meningitidis.

XX

PN WO200003003-A2.

XX

PD 20-JAN-2000.

XX

PF 09-JUL-1999; 99WO-GB02205.

XX

PR 10-JUL-1998; 98GB-0014902.

XX

PA (UYNO-) UNIV NOTTINGHAM.

XX

PI Ala'Aldeen D, Todd I;

XX

XX WPI; 2000-147612/13.

DR N-PSDB; AAZ49702.

XX

PT Generation of cell lines and clones specific to a particular protein
 PT for screening antigenic peptides which are used as vaccines in treating
 XX meningococcal, gonococcal infections

Claim 59; Page 42-45; Sipp; English.

CC The present sequence is N. meningitidis (strain SD, serogroup B
 CC (B:15:P1.16)) T-cell stimulating protein A (TspA). TspA is a CD4+ T-cell
 CC stimulant. It can be produced recombinantly using lambda ZapII phage
 CC library comprising the DNA encoding TspA. TspA has antibacterial and
 CC anti-inflammatory activity and can be used in vaccine formulations
 CC against neisserial diseases like meningitis, gonorrhoea, septicemia,
 CC septic arthritis and pelvic inflammatory diseases. T-cell lines and
 CC clones specific to neisserial proteins can be generated for screening
 CC meningococcal or gonococcal genomic phage display libraries to
 CC identify peptides which stimulate T-cell lines and clones.

XX Sequence 880 AA;

Query Match 100.0%; Score 4404; DB 21; Length 880;
 Best Local Similarity 100.0%; Pred. No. 3.5e-255;
 Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGLPRRCPMWTKFTDCTRSNRITOPTHRGYIILKNNRQIKLIAASVAASFOAHAGL 60
 DB 1 MPAGLPRRCPMWTKFTDCTRSNRITOPTHRGYIILKNNRQIKLIAASVAASFOAHAGL 60
 QY 61 GGLNIQSNLDPFSGSIIVTGEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAYSSQQA 120
 DB 61 GGLNIQSNLDPFSGSIIVTGEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAYSSQQA 120
 QY 121 VRDPLVFRIGAGAQVREYTAIILDPVGYSPKTSALS DGKTHRTAPTAEQENQAKAL 180
 DB 121 VRDPLVFRIGAGAQVREYTAIILDPVGYSPKTSALS DGKTHRTAPTAEQENQAKAL 180
 QY 181 RKTOKOSANAAPVPAKNGKTHTVKGETVQVQIAAALRPKHLTLTLEQVADALLKANPNVSA 240
 DB 181 RKTOKOSANAAPVPAKNGKTHTVKGETVQVQIAAALRPKHLTLTLEQVADALLKANPNVSA 240
 QY 241 HGLRAGSVLHILPNLRKAEQPKPQAKPAEATASMPSEPSKQATVEKPEKPEAKVAA 300
 DB 241 HGLRAGSVLHILPNLRKAEQPKPQAKPAEATASMPSEPSKQATVEKPEKPEAKVAA 300
 QY 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAAGALDPTDETGNVSEPEVQVSAE 360
 DB 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAAGALDPTDETGNVSEPEVQVSAE 360
 QY 361 EETESGLFGGSYTLILLAGGGAALIALILILRLAQSKARTEERSVPEEPDLDAAADGI 420
 DB 361 EETESGLFGGSYTLILLAGGGAALIALILILRLAQSKARTEERSVPEEPDLDAAADGI 420
 QY 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEELSAKQTFDVTETPSNRIDLDFDLSA 480
 DB 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEELSAKQTFDVTETPSNRIDLDFDLSA 480
 QY 481 AAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNPVEIVDTPEPESVAQ 540
 DB 481 AAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNPVEIVDTPEPESVAQ 540
 QY 541 TANKPPTVDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
 DB 541 TANKPPTVDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
 QY 601 TPEELHDFLKVYETDAVAETAPETPDFAAADDLSALLOPAEAPSVBEENTETVAETPDF 660
 DB 601 TPEELHDFLKVYETDAVAETAPETPDFAAADDLSALLOPAEAPSVBEENTETVAETPDF 660
 QY 661 NATADLSALLOQSEVPVPAVEENAAEIVADDLSSALLOPAEAPVEENVETVAETSDFHTA 720
 DB 661 NATADLSALLOQSEVPVPAVEENAAEIVADDLSSALLOPAEAPVEENVETVAETSDFHTA 720
 QY 721 ADLSALLOQAEVPAVEENVTKTVAETPDENATADDLSALLOQSEVPVPAVEENAAETITLET 780

DB 721 ADDLSALLOQAEVPAVEENVTKTVAETPDENATADDLSALLOQSEVPVPAVEENAAETITLET 780
 QY 781 PDSNTSEADALPDFLKGDEETVDSIYLSSENI PNADTSPBSVSGSDAPSEAKYDLA 840
 DB 781 PDSNTSEADALPDFLKGDEETVDSIYLSSENI PNADTSPBSVSGSDAPSEAKYDLA 840
 QY 841 EMYLEIGDRDAAAETVQKLLLEAEAGDVLKRAQALAEGLI 880
 DB 841 EMYLEIGDRDAAAETVQKLLLEAEAGDVLKRAQALAEGLI 880

RESULT 2
 ABP78566

ID ABP78566 standard; Protein; 244 AA.

XX AC ABP78566;

XX DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 3662.

XX KW Antibacterial; infection; vaccine; gene therapy.

XX OS Neisseria gonorrhoeae.

PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX DR N-PSDB; ABZ39536.

XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 medicament for treating or preventing N. gonorrhoeae infection

XX PS Disclosure; Page 453; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.

XX SQ Sequence 244 AA;

Query Match 25.5%; Score 1122.5; DB 24; Length 244;
 Best Local Similarity 64.6%; Pred. No. 1.5e-59;
 Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;

QY 514 VYEPETFNPNPVEIVDTPEPESVQATKENTKPTVDTDFSDNLPNNHIGTETASAKP 573
 DB 1 VYEPETFNPNPVEIVDTPEPESVQATKENTKPTVDTDFSDNLPNNHIGTETASAKP 60
 QY 574 ASPSGLAGFLKASSPETILEKTVAEVQTPPELHDFLKVYETDAVAETAPETPDFNAAAD 633
 DB 61 AAPSGLAGFLKASSPETILEKTVAEVQTPPELHDFLKVYETGAETAETPTPDFNAAAD 120
 QY 634 LSALLOPAEAPSVBEENITETVAETPDFNATADDLSALLOQSEVPVPAVEENAAEIVADDL 693
 DB 121 LSALLOPAE----- 130
 QY 694 LLQPAEAPAEVENVETVAETSDFHTAADDLSALLOQAEVPAVEENVTKTVAETPDENAT 753

Qy	634	LSALLQPAEAPSVENITETVAETPDFNATADALLQPPSEVPAVENAAEIVADDLSA	693
Db	121	LSALLQPAEA-----	130
Qy	694	LLQPAEAPAVEENVTVTAETSDFTAAADLSALLQPAEVPAAENVTKTVAEIPDFNAT	753
Db	131	-----PAVEENAAEITLET-----	130
Qy	754	ADDLSALLQPPSEVPAVENAAEITLETSDNTSEADALPDFLKDGEEETVDWSIYLSEN	813
Db	131	-----PAVEENAAEITLETSDNTSEADALPDFLKDGEEETVDWSIYLSEN	177
Qy	814	IPNNADTSFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEIVOKLLEAEAGDVLKRAQA	873
Db	178	IPNNADTCFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEIVOKLLEAEAGDVLKRAQA	237
Qy	874	LAQELGI 880	
Db	238	LAQELGI 244	

RESULT 4
ABP78509 standard; Protein; 275 AA.

XX	AC	ABP78509;
XX	DT	07-MAR-2003 (first entry)
XX	DE	N. gonorrhoeae amino acid sequence SEQ ID 3548.
XX	KW	Antibacterial; infection; vaccine; gene therapy.
XX	OS	Neisseria gonorrhoeae.
XX	PN	WO200279243-A2.
XX	PD	10-OCT-2002.
XX	PF	12-FEB-2002; 2002WO-IB02069.
XX	PR	12-FEB-2001; 2001GB-0003424.
XX	PA	(CHIR-) CHIRON SPA.
XX	PI	Fontana MR, Pizza M, Massignani V, Monaci E;
XX	DR	WPI; 2003-058415/05.
XX	DR	N-PSDB; ABZ39479.
XX	PT	New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX	PT	medicament for treating or preventing N. gonorrhoeae infection -
XX	PS	Disclosure; Page 447; 815pp; English.
XX	CC	The present invention relates to proteins from Neisseria gonorrhoeae.
XX	CC	Also disclosed are the nucleic acid molecules encoding the proteins and
XX	CC	antibodies that specifically bind to the proteins. The composition
XX	CC	comprising the protein, nucleic acid or antibody is useful for the
XX	CC	manufacture of a medicament for treating or preventing N. gonorrhoeae
XX	CC	infection, this may be in the form of a vaccine or gene therapy.
XX	CC	Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX	CC	molecules of the invention.
XX	SQ	Sequence 275 AA;

Query Match 23.3%; Score 1025; DB 24; Length 275;
Best Local Similarity 83.4%; Pred. No. 1.2e-53;
Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;

Db	131	-----	130
Qy	754	ADDLSALLQPPSEVPAVENAAEITLETSDNTSEADALPDFLKDGEEETVDWSIYLSEN	813
Db	131	-----PAVEENAAEITLETSDNTSEADALPDFLKDGEEETVDWSIYLSEN	177
Qy	814	IPNNADTSFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEIVOKLLEAEAGDVLKRAQA	873
Db	178	IPNNADTCFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEIVOKLLEAEAGDVLKRAQA	237
Qy	874	LAQELGI 880	
Db	238	LAQELGI 244	

RESULT 3
ABP80077 standard; Protein; 244 AA.

XX	AC	ABP80077;
XX	DT	07-MAR-2003 (first entry)
XX	DE	N. gonorrhoeae amino acid sequence SEQ ID 6684.
XX	KW	Antibacterial; infection; vaccine; gene therapy.
XX	OS	Neisseria gonorrhoeae.
XX	PN	WO200279243-A2.
XX	PD	10-OCT-2002.
XX	PF	12-FEB-2002; 2002WO-IB02069.
XX	PR	12-FEB-2001; 2001GB-0003424.
XX	PA	(CHIR-) CHIRON SPA.
XX	PI	Fontana MR, Pizza M, Massignani V, Monaci E;
XX	DR	WPI; 2003-058415/05.
XX	DR	N-PSDB; ABZ41047.
XX	PT	New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX	PT	medicament for treating or preventing N. gonorrhoeae infection -
XX	PS	Disclosure; Page 674; 815pp; English.
XX	CC	The present invention relates to proteins from Neisseria gonorrhoeae.
XX	CC	Also disclosed are the nucleic acid molecules encoding the proteins and
XX	CC	antibodies that specifically bind to the proteins. The composition
XX	CC	comprising the protein, nucleic acid or antibody is useful for the
XX	CC	manufacture of a medicament for treating or preventing N. gonorrhoeae
XX	CC	infection, this may be in the form of a vaccine or gene therapy.
XX	CC	Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX	CC	molecules of the invention.
XX	SQ	Sequence 244 AA;

Query Match 25.5%; Score 1122.5; DB 24; Length 244;
Best Local Similarity 64.6%; Pred. No. 1.5e-59;
Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;

Qy	514	VYEPETPNPNVEIVIDTPEPSVAQTAEKPEVDTDFSDNLPNNHIGTETASAKP	573
Db	1	VYEPETPNPNVEIVIDTPEPSVAQTAEKPEVDTDFSDNLPNNHIGTETASAKP	60
Qy	574	ASPSGLAGFLKASSPETILEKTVAEVOTPELHDFLKVYETDAVAETAPETPDFAAAD	633
Db	61	ASPSGLAGFLKASSPETILEKTVAEVOTPELHDFLKVYETDAVAETAPETPDFAAAD	120

QY 167 PTAESQENQAKALRKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQ 226
 DB 61 PKAESQENQAKALRKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQ 120
 QY 227 VADALLKANPNVSAHGRLEAGSVLHHPNLNR-----IKAEQPKQTAAPKAETASMP 278
 DB 121 VADVLLKANPNVSAHGRLEAGSVLHHPNLNR-----IKAEQPKQTAAPKAETASMP 180
 QY 279 SEPSKQAT-----VEKPEKPEAKVAAPKAEPKPAVRPE-----313
 DB 181 SEPSKQATVEKPEKPEAKVAAPKAEPKPAVRPE-----313
 QY 314 -PVPAAANTAASATAESAP 331
 DB 241 GPVPAANTAASATAESAP 259

RESULT 5

ABP79684
 ID ABP79684 standard; Protein; 275 AA.

AC ABP79684;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 5898.
 KW Antibacterial; infection; vaccine; gene therapy.
 OS Neisseria gonorrhoeae.

XX WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Massignani V, Monaci E;
 DR WPI; 2003-058415/05.
 DR N-PSDB; ABZ40654.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 PS Disclosure; Page 619; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.

XX Sequence 275 AA;

Query Match 23.3%; Score 1025; DB 24; Length 275;
 Best Local Similarity 83.4%; Pred. No. 1.2e-53;
 Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;

QY 107 LGDKAVTAVSSEQAVRPVLFRIAGAGQVREYTAILDVGVGSPKTSALSDGKTRKTA 166
 DB 1 LGDKAVTAVSQAQAVRPVLFRIAGAGQVREYTAILDVGVGSPKTSALSDGKTRKTA 60
 QY 167 PTAESQENQAKALRKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQ 226

DB 61 PKAESQENQAKALRKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQ 120
 QY 227 VADALLKANPNVSAHGRLEAGSVLHHPNLNR-----IKAEQPKQTAAPKAETASMP 278
 DB 121 VADVLLKANPNVSAHGRLEAGSVLHHPNLNR-----IKAEQPKQTAAPKAETASMP 180
 QY 279 SEPSKQAT-----VEKPEKPEAKVAAPKAEPKPAVRPE-----313
 DB 181 SEPSKQATVEKPEKPEAKVAAPKAEPKPAVRPE-----313
 QY 314 -PVPAAANTAASATAESAP 331
 DB 241 GPVPAANTAASATAESAP 259

RESULT 6

ABP78538
 ID ABP78538 standard; Protein; 253 AA.

AC ABP78538;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 3606.
 KW Antibacterial; infection; vaccine; gene therapy.
 OS Neisseria gonorrhoeae.

XX WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Massignani V, Monaci E;
 DR WPI; 2003-058415/05.
 DR N-PSDB; ABZ39508.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 PS Disclosure; Page 450; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.

XX Sequence 253 AA;

Query Match 18.7%; Score 821.5; DB 24; Length 253;
 Best Local Similarity 66.8%; Pred. No. 1.6e-41;
 Matches 181; Conservative 11; Mismatches 40; Indels 39; Gaps 5;

QY 255 LNKIKAEQPKQTAAPKAETASMPSEPSKQATVEKPEKPEAKVAAPKAEPKPAVRPEP 314
 DB 1 LKNLKQKLEPP-----KQKRNRPD-ENPNPQCCKLRLRRQNANPDL 42
 QY 315 VPAANTAASATAESAPQEAASAIPTDPTDCTGNVS-----EPVEQVS 358
 DB 43 YFQIILPHRKPLNPNPPQEAASAIPTDPTDCTGNVS-----EPVEQVS 102
 QY 359 ABEETES----GLFGGSYTLILLAGGAALTA-LLLLRLAQSKRARTTESVPEEPDLD 413

Db 103 AEEETESGLFDGLFGGSYTLILLACGGTALLALLLLLRLLAQSKRARTESVPEEPDLD 162
Qy 414 DAADGGIETFAEVETPATEPAPKNDVNDTLALDGESEBELSAKQTFDVTETPSNRID 473
Db 163 DAADGGIKITFAEVETPATEPAPKNDVNDTLALDGESEBELSAKQTFDVTETPSNRID 222
Qy 474 LDFDSLAAQNGILSGALTQDEETQKADAD 504
Db 223 LDFDSLAAQNGILSGALTQDEETQKADAD 253

RESULT 7
ABP79950
ID ABP79950 standard; Protein; 132 AA.
XX AC ABP79950;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 6430.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX FN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.
XX PR 12-FEB-2001; 2001GB-0003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; ABZ40920.

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 656; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX Sequence 132 AA;

Qy Query Match 14.1%; Score 620.5; DB 24; Length 132;
Best Local Similarity 97.0%; Pred. No. 7.4e-30;
Matches 128; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
Qy 374 LLLAGGGAALIA-LLLLLRLAQSKRARTESVPEEPDLDAADDGIETFAEVETPAT 432
Db 1 LLLAGGGAALIA-LLLLLRLAQSKRARTESVPEEPDLDAADDGIETFAEVETPAT 60
Qy 433 PEPAPKNDVNDTLALDGESEBELSAKQTFDVTETPSNRIDLDFDSLAAQNGILSGALT 492
Db 61 PEPAPKNDVNDTLALDGESEBELSAKQTFDVTETPSNRIDLDFDSLAAQNGILSGALT 120
Qy 493 QDEETQKADAD 504
Db 121 QDEETQKADAD 132

RESULT 8
ABJ26247
ID ABJ26247 standard; Protein; 1095 AA.
XX AC ABJ26247;
XX DT 17-APR-2003 (first entry)
XX DE Aspergillus fumigatus essential gene protein #905.
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response.
XX OS Aspergillus fumigatus.
XX FN WO200286090-A2.
XX PD 31-OCT-2002.
XX PF 23-APR-2002; 2002WO-US13142.
XX PR 23-APR-2001; 2001US-285697P.
XX PR 27-APR-2001; 2001US-287066P.
XX PR 05-JUN-2001; 2001US-295890P.
XX PR 09-JUL-2001; 2001US-303899P.
XX PR 31-AUG-2001; 2001US-316362P.
XX FA (ELIT-) ELITRA PHARM INC.
XX PI Jiang B, Tishkoff D, Zamudio C, Broshkin AM, Hu W, Lemieux SM;
XX DR WPI; 2003-093124/08.
XX PT New purified or isolated nucleic acids of essential genes of
XX PT Aspergillus fumigatus, useful for treating or preventing infections by
XX PT A. fumigatus, or for treating a non-infectious disease in a subject
XX PT e.g. cancer -
XX PS Disclosure; Page -; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This sequence represents a protein of one of the essential genes
XX of Aspergillus fumigatus of the invention.

XX Sequence 1095 AA;

Query Match 8.1%; Score 358; DB 24; Length 1095;
Best Local Similarity 22.1%; Pred. No. 6.1e-13;
Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;

Db 602 TVBEIVAPSVVSVAPSVESVEENVEESVAENVEESV----- 647
QY 584 KASSPETILEKTVAEQTPEELHDFLKVYETDAVETAPPTDFNAAADLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVBEIVAPSVVSVAPSVESVEEN-VEESVAENVEESVA 704
QY 644 PSVEENTETVAETPDNATADLSALLQSPVAPVEENAAEIVADDSALLQPAEAPAV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPSVVSVAPSVESV 759
QY 704 EENVETVAETSDPHTAADDLSALLQPAEVPAPVEENVTKVAETPDNATADLSALLQ 763
Db 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEII-----VAPSVEE 805
QY 764 SEVPAVEENAAEITLETDPDNTSADALPFLKDGREETVDWMSIYLSEENIPNADTSFP 823
Db 806 SVAPSVESVA---ENVATNLSD-NLLSNLLGGIETEIEKDSILNIEEVKENVVTI- 859
QY 824 SESVGSAPSEAKYDLAEMYLEIGRDAARAEVTKLLEAEAGDVLKRA 871
Db 860 LENV-BETTAESVTTFSNILEEIQENTITNTDIEKLEELHENVLSAA 906

RESULT 12
ABU25647
ID ABU25647 standard; Protein; 1000 AA.

AC ABU25647;
DT 16-APR-2003 (first entry)

Aspergillus fumigatus essential gene protein #305.

Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
cancer; contamination; biofilm; antibody; immune response.

Aspergillus fumigatus.

WO200286090-A2.

31-OCT-2002.

23-APR-2002; 2002WO-US13142.

23-APR-2001; 2001US-285697P.

27-APR-2001; 2001US-287066P.

05-JUN-2001; 2001US-295890P.

09-JUL-2001; 2001US-303899P.

31-AUG-2001; 2001US-316362P.

(ELIT-) ELITRA PHARM INC.

Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

WPI; 2003-093124/08.

New purified or isolated nucleic acids of essential genes of

Aspergillus fumigatus, useful for treating or preventing infections by

A. fumigatus, or for treating a non-infectious disease in a subject

e.g. cancer

Disclosure; Page -; 175pp; English.

The invention relates to novel purified or isolated nucleic acids of
essential genes of Aspergillus fumigatus. The isolated nucleic acids of
the invention are used to treat or prevent infections by a pathogenic
organism such as A. fumigatus, to treat a non-infectious disease in a
subject (e.g. cancer), to prevent or contain contamination of an object
by A. fumigatus, or to prevent or inhibit formation on a surface of a
biofilm comprising A. fumigatus. The polynucleotides are useful for
expressing recombinant protein for characterisation, screening or
therapeutic use, as markers for host tissues in which the pathogenic
organisms invade or reside, for comparing with the DNA sequence of A.

CC fumigatus to identify duplicated genes or paralogues having the same or
similar biochemical activity and/or function, for comparing with DNA
sequences of other related or distant pathogenic organisms to identify
potential orthologous essential or virulence genes, for selecting and
making oligomers for attachment to a nucleic acid array for examination
of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.

XX SQ Sequence 1000 AA;

Query Match 7.2%; Score 315; DB 24; Length 1000;

Best Local Similarity 24.2%; Pred. No. 2e-10;

Matches 194; Conservative 109; Mismatches 289; Indels 210; Gaps 41;

QY 101 TAKVHKLGDKAVIATVSSEQAVRDLVFRIGAGQVREYTAIDLPVGVSPKTKALSDBGK 160
Db 29 SAELDKV--EAKEAAARQAEDAKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 86
QY 161 THRKTAPTAEQENONAKALRKTDKDSANAAVKPAYNGKTHTVRKGETVQIAAARPK 220
Db 87 DNEEAP-ACGDPPTNESAEEDVTKPTDENS---PAPDGA----- 123
QY 221 HLTLEQVADALLKANPNVSAHGRLRAGSVLHPIMLNRIKAEQPKP--TAKPKAETASMP 278
Db 124 ---EESGSDPPEAEQVAE-----VQVPT-----EQPAPKPAPEPSSEAPAP 165
QY 279 SEPSKQATVEKPV--EKPEAKVAAPAE-----KA-EKPAVTRPEV----- 315
Db 166 TETAENNTSEAVAPAGATADIAAETTKESTDEKAPPEEPAAESPVEDKQDEATKALA 225
QY 316 -----PAANTA-ASETAESAPQEA-----ASAITDPTDE-TGNAVESEPQV 357
Db 226 AEEGKTTPAADAEEVADEAAEAPAEASSTDEVAQPEATPAEATPEEKEKEDAPPEPQP 285
QY 358 SAEETESGLFGSVYLLLAGGGAALIALLLLAQSKRAARRTEESVPEEPDLDDAAD 417
Db 286 LTEEVVE-----SAPSVESAEPAEAPAE-----SAPSVESAEPAEAPAE 318
QY 418 DGIETITFAEVTATP---EPAPKND--VNDTLALDGESEBELSAKQTFDVTETPSNRI 472
Db 319 E--EPPAAESAPAVPEVEEAPQDETPVQKAPAAEQSTAEDVTPEEPAPVEE----- 369
QY 473 DLDFTSLAAQNGILSGALTODEET--QKRADADWNATSTDSVVEPETFPYNPVEIVI 530
Db 370 -----ASAKEPV-----ABEPTFGDKSPVEEAEEAPPAAESAESESTPVEEA- 415
QY 531 DTPEPESVAQ---TAENKPEVTD-----TDFSNLPSNNHIGTETASAKPASPSGLAG 581
Db 416 -APTESAAESAPAEAEATESISVCEPAAADAEGE-----REESAHEEPVDE--- 464
QY 582 FLKASSPPTILEKTVAE---VQTPPELHDFLKVYETDAVETAP---ETPDNAAADDLS 635
Db 465 -APAEAPAVPEEPAEAESESPVEDPAPVEEAPAEAPAAAAEPIETEEPAEAEVPAEPA 523
QY 636 ALLQPAEAPS-VEENITETVA-----ETPDFNATADLSALLQSPVAPVEENAAEIVADD 690
Db 524 PQOEPAEEPTPVEKSVPEESAPAEERIIVEEAPAEEST-----PAPPAVPEEVPABE- 575
QY 691 LSALLQPA--EAPAVEE-NVETV-AETSDPHTAADDLSALLQPAEV-PA-VEENVTKIV 744
Db 576 -SAPVEAPVPEEPAEAEAPPTESAREEEAPMEEATEESASVEPASVEPAPVEPVEEP 634
QY 745 ABIPDFNATADLSALLQSPSEVPAPVEENAAETITLETPTDNTSEADALPFLKDGREETVD 804

Db 635 A-----AAAREI-----PABEPAVVEGAPVEESVPVKEAAPVDA-----EFSEETPD 677

QY 805 WSVLSEENIPNNAADTFSPSES 826

Db 678 PT---PVIEISRDIDALEPLEA 696

RESULT 13

AAAR27745

ID AAR27745 standard; Protein; 1822 AA.

XX AC AAR27745;

XX DT 25-MAR-2003 (updated)

XX DT 04-MAR-1993 (first entry)

XX DE Extracellular factor related protein.

XX KW EF; detection; prevention; screening; diagnostic.

XX OS Streptococcus suis type II (non-pathogenic).

XX FH Key Location/Qualifiers

FT Peptide 1..46

FT Peptide /note= "signal peptide"

FT Peptide 47..1822

FT Region /note= "mature peptide"

FT Region 858..861

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 934..937

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 999..1002

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1075..1078

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1264..1267

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1362..1365

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1438..1441

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1514..1517

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1590..1593

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1666..1669

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1741..1744

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FN W09216630-A1.

XX 01-OCT-1992.

XX 19-MAR-1992; 92WO-NL00054.

XX 21-MAR-1991; 91NL-0000510.

XX (DIER-) CENT DIERGENESKUNDIG INST.

XX Smith HE, Vecht U;

XX WPI; 1992-349215/42.

XX N-PSDB; AAQ29471.

XX Deoxyribonucleic acid encoding virulence characteristic of

XX Streptococcus suis - useful for antibody and polypeptide for

XX diagnosing and preventing infections in pigs and humans

XX Claim 9; Fig 1b; 86pp; English.

XX The sequence is that of the extracellular factor related protein

XX from Streptococcus suis type II (non-pathogenic) which allows the

CC detection and the prevention of infections by S. suis in a more

CC effective manner than was previously possible. It facilitates

CC screening of e.g. pigs and elimination of infected and carrier pigs

CC can then be carried out. The new diagnostic tests can distinguish

CC between avirulent and virulent strains. It may be used in the prodn.

CC of a vaccine. See also AAR27744 and AAR27746.

CC (Updated on 25-MAR-2003 to correct PN field.)

XX

SQ Sequence 1822 AA;

Query Match 7.1%; Score 313; DB 13; Length 1822;

Best Local Similarity 20.3%; Pred. No. 5.8e-10;

Matches 209; Conservative 163; Mismatches 356; Indels 302; Gaps 46;

QY 37 NNRQIKLIAASV-----AVAAAFQAHAGLGLNIGLNISLD--BPFSGSITVTGEEAKAL 87

Db 682 DNERLKLGLPDSAFVTNSDGTVSVDYSA---GGVNDGATDIKNATTNLADTRNEAKA- 736

QY 88 LGGGSVTVSE--KGLTAK---VHKLGDKAVIAVSSEQAVRDPVLVFRIGAGAOVREYTA 141

Db 737 --GIDTKLAHEHKALEAKRDEAFSKIDDDISLSAEQQAQAKNAVA---AAGDALKE--- 788

QY 142 ILDPVGYSPK-----TKSALSOGKTHRKTPAPTAEQONQAKALRKTDKXSANAAY- 193

Db 789 -LDNKATEAKEKIDKATTASEINDAKTNGEI--NLDSAEAVGEKAINQAKELAKAEVE 845

QY 194 -----KPAY-----NGKTHTVRK---GETVKQIAAAIRPKHLT-- 223

Db 846 NKAFEALEKVNNNPNLLBEEKKAYFDDIKESKEVAVEKINNAENTAETITAAIDBAEIAYN 905

QY 224 -----LEOVAD---ALLKANPNVSAHGLRAGSVLHILPNLNRIKAEQPKPQ 266

Db 906 EDVINAQAQDALANKLEKDSBETKAAIDANPNLTPEEKAKA-----IAKVELVNNAESDI 960

QY 267 TAPKAEATASMPSEPSKOATVEKPEKAAVAPEAK---AEKPAVRPEPVPAANTAAS 323

Db 961 LSKPTPTETQAVEDKA-----DKDLAKVELQAADGAKKGIENPNLTPEEKDVAKKAVE 1015

QY 324 ETAESAPOEAAASADID---TPTDETGNVSEPEVQESAAEE--ETESGLFGGSGVTLILAG 378

Db 1016 DAV-----KVATDAIDKASTPT-EVDTATSDGVKAIADAEFKATQKD----- 1056

QY 379 GGAALIALLLLLRLAQSRAKRRTTES-----VPEEPDLDDAADGIEITFAEVE 428

Db 1057 -----AKNKIAKEAESAKAIDDPNLTPEEKESAKNAVEAAKAVATAAID 1102

QY 429 TPATPEPAPKNDVNDTLALDGESEELSASAKQTFDVTETDP-----SNRIDLDFDSLAA 481

Db 1103 KASTPD-----AVQVEEDKGVAAINLITAKADAKGVIAAKLADEIKKLEDKQAE 1151

QY 482 AQNGILSGALTQDEET-----OKRADADWNAIESDTSVYEPETTFNPNVPI-- 528

Db 1152 AEKAIASMTNNEEKAIKAKALQDVVDKGALEDAARVATNBIHEAITTEKAKAELAG 1211

QY 529 ---VIDT-PEPESVAQTAENKP-----ETVDTDFSDNLPNNHIGTEE 567

Db 1212 EKSLTDTGKEARDAVELAKDKELGKEAIRTEEEAEKIVKEKLAEDTRKAIEDPNLSDED 1271

QY 568 TASAKPASPGLAGFLKASSPETILEKTVAEVQTPPEELHDFLKVYETDAVAETAPETPDF 627

Db 1272 KQAEIKKLTDAVAKTLATWRDNA--DKRTQEAekaQALADLEKAKETQKIAD----- 1321

QY 628 NAAADDLSALLOPAEAPSVEEN-----ITETVAETPPFN----- 661

Db 1322 KKAIDRLTILVKDGELEATKQDAKTKIAKDAKAAKAIASNPMLTDAEKKTTFTDAVDAEV 1381

QY 662 ATADD--LSALLOPSEVPVAVEANAEIVADD--LSALLQPA-----EAPAVEE-----NV 707

Db 1382 AKANDAI SAATSPADVQKEEDAGVAIAEDVLDAKQDAKKNIAKDAKAAKAEIGSNPNL 1441

QY 708 TE-----TVAETSDPHTAADLSALLQPAEVPVAVEENTKTVVAEIPDFPNATADDLSALL 761

Db 1442 TDAEKKTTFTDAVDAEVAEKANDAI SAATSPADVQKEEDAGVAIAE-----DVLDAAK 1493

QY 762 OPSEVPAVENAAEITLTETPDNTSEADALPDFLKOCGEETVDSIYSEENIPNADTS 821
 Db 1494 QDAKNKIAGES-----DAKSAIDANFN-LTDAEKSSAKAV-----DADAK 1534
 QY 822 FPSESV-GSDAPSEAKY-----DLAEMYLEIGDRDAAAETVQKL-----LE 861
 Db 1535 AATDAIDASTPVEAQAEDKGVGSIAQDVLDAKQDAKNKIAKEVAKAAEIDANFNLS 1594
 QY 862 EAEGDVLKRA 871
 Db 1595 DAEKEASKA 1604

RESULT 14

ABB62991
 ID ABB62991 standard; Protein; 1468 AA.

AC ABB62991;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 15765.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07094.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

XX Disclosure; SEQ ID NO 15765; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 sequences (AB101840-AB116175) and the encoded proteins
 (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1468 AA;

Query Match 6.6%; Score 289.5; DB 22; Length 1468;

Best Local Similarity 22.5%; Pred No. 1.1e-08;

Matches 207; Conservative 111; Mismatches 337; Indels 266; Gaps 42;

QY 65 IQSNL-----DFFPGSITVTGEEAKALGGSVTVSEKGLTAKVHKLGDKA----- 111

Db 366 IQTLQLEDGEP-ETSTQVADEEL-----IPVNE-GETESVTKAADETEIPETP 415

QY 112 -----VIAVSEQAV-----RDPVLVPRIGAGQVREYTAILDPVGYSPKTKSA 155
 Db 416 EDIASGHELDVAAASHDDDEVETPKQEEPTALKIBQPAQVESQNVDEVGPEEPQAE 475
 QY 156 LSDGKTHRTAPTAEQONONAKALRKTOKSANAARKPAYNGKTHVTKGTVKQIAA 215
 Db 476 MEIEKEKEETAAT-----EAAEVELPATEAKPDMESEIDQVEQ-SA 515
 QY 216 AIRPKHLTLEQVADALLKANPNVSAHGRLRAGSVLHPNLRIRKAEQPKQAKPKAETA 275
 Db 516 DIXP-----TSGEDDLNEAN-----QLNQDQVTEPAALQEIEEKEPEQVVDMHEDA 563
 QY 276 SMPSEPSKOATVEKPV-----EKPEAKVAA-----PEAKAEKPAV 310
 Db 564 D-----EATTEKPVQDHQDEVEALTEAEHGQLTYTNDDELAPVIEPELTTEPEI 616
 QY 311 R-----PEVPAANTAASETAASAPOEAAASAIIDTPTDGTGNAVSPVE 355
 Db 617 AENTTELPEQOQQOQDEEIEEAAALAEHQVQVPEE-NANAIKTP-EVSEAIAP-- 671
 QY 356 QVSAEBETESGLFGGSYTLILAGGGAALIALILLRLAOSKRARRTEESVPEEPDLDDA 415
 Db 672 --QADLETPSA-----TEKEI--BPQTADHV 693
 QY 416 ADDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEELSAKOTFDVETDTPSNRIDLD 475
 Db 694 EEPVATERIQAEPEPEIETPNDEVSQVAD-QAEPPIKSDADIAEAAEE----- 747
 QY 476 FDSLAAQNGILSGALTQDEETQKRADADWNAIESTDSVYEPETFNPNVVEIVIDTPEP 535
 Db 748 ----PAQTEVAPEVETETEAESQAQTSPEISVEQ-----EVAIDKNN 787
 QY 536 ESWAQTAKNKEPTVDTDFSNLPSNNHIGTEETASAKPASPGLAGFLKASSPE---TIL 592
 Db 788 DNIIQGAEPNEDELEIPEPQPMPEEELPVQOEASAVPEDH-----KLTKEPMDYTSV 840
 QY 593 EKTVAEVOTPEELHDFLKVYETDAVET-APETPDFNAADDLSALLQPAEAPSVEENIT 651
 Db 841 EH-LQELSRPEETAAIEAESDSRESTLAPEADAFIAI--VSEERDPEDHP---QNFH 894
 QY 652 ETVAETPDFNATADDLSALL-----QPSE-VPAVE---ENAAEIVADDLSA 693
 Db 895 ES-----ESIADILSLMLEADSTVPPVPFGQPAQTPVMVEANGSENEEQAEHAISD 947
 QY 694 LLQPAEAPAVEENVETVAETSDHTAADDL--SALLQPAEYPAVENVTVAIEIDFN 751
 Db 948 L--QAEADETQEH-AEQVQVNEPETAPELFAEAKDEEDVPALPES----- 992
 QY 752 ATADDLSALLQPSVPAVEENAAEITLTETPDNTSEADALPDFLKOCGEETVDSIYSE 811
 Db 993 ----DESKTQPSAL-EVEQHIASEAQTPVEDHGYDE-EHQDQDEHQDEDEPLKE 1046
 QY 812 ---ENIENNADTSF-----PSESVGSDAPSEAKYDLAE-----MYLEIGDRDAAAETVQKL 859
 Db 1047 VETESLPEEATTEKFORPEPEEVNTPLEELPDLTDHSDVNRNRELQESLILESTNI 1106
 QY 860 --LEEAGDVLKRAQALAE 878
 Db 1107 DGLQELDNHIEEPEAAHAI 1127

RESULT 15

ABB68110

ID ABB68110 standard; Protein; 753 AA.

XX AC ABB68110;

XX AC ABB68110;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 31122.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

[illegible]

Search completed: December 12, 2003, 17:30:57
Job time : 50 secs

```

Query Match      6.5%; Score 287.5; DB 22; Length 753;
Best Local Similarity 21.9%; Pred. No. 6.2e-09;
Matches 191; Conservative 109; Mismatches 331; Indels 243; Gaps 40;

QY      63 LNIQSNLDPEFGSGIITVTGFEAKALLGGGVTSEKGLTAKVHKLGDKAVIASVSEQAVR 122
Db      32 IDVQKTDAP-----AVNGDAATPKEGGDDEAAAVEKKETEE--HSENDKDUITTEKSAVAE 85

QY      123 DPLVFRIGAGAOVREYTAILDPVGYGPKTKSALS DGKTKRKTAPTAAESOENOKAKALRK 182
Db      86 GGDVAETAKEE-----GSPKEAAGED-----ITPLAD-----ESIKSKSK 123

QY      183 TD-----KKDSANAARKPAYNGKTHTVRKGETVTKQIAAAIRPKHLTLBQVA 228
Db      124 KDKVKKKWSFRSISFGKKDKQ-----KPAKEEATSTPTSGTSTPTTAA-----EAAPAG 173

QY      229 DALLKANPNYSAHGRLRAGSVLHPIPNLNRITKARQPKQTA KPAAETASMPSESKQATVE 288
Db      174 DAUV-AEPSVATNG-----EAKRP-----AETATATSEFASKD--E 206

QY      289 KPVEKPEAKVAAPAEAKEPFAVRPEPVPAANTAASETAAESAPOEAAASAIDTPTDFTGN 348
Db      207 KPAENGSAETEQKQANGETEKAAPAP-----SVVEEAAKPKPAEEPA TV--TATESNTT 258

QY      349 AVSE-FVEQVSAABEETESGTFGGSYTLLLAGGGAALLALLLLRLAQSKARTEESVPE 407
Db      259 ATEVPVKESQPEPEVVTNGHGAEA--LTNGSSNGLA-----ESPVTE 300

QY      408 EEPDLDDAADGIGIITFAEVETPATPEAPKNDVNDT-----LALDGSEBELSAKQT 460
Db      301 TAPVADN-----TPSNVDDPEPHQNGTNGTTTTPPTPVATEIKGQOIEASSE 348

QY      461 FDVETDTPSNRIDLDFDSLAAQNGILSGALTQDEETQKRAADWNNAIESTSDSYPEPTF 520

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:30:02 ; Search time 22 Seconds
(without alignments)
1692.434 Million cell updates/sec

Title: US-09-743-674-2
Perfect score: 4404
Sequence: 1 MPAGRLPRCPMMTKFTDCT.....EEAECDVVKRAQALQELGI 880

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles!.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	398.5	9.0	975	4	US-09-252-991A-28749 Sequence 28749, A
2	315.5	7.2	1786	3	US-08-973-462-8 Sequence 8, Appli
3	289	6.6	8991	4	US-08-714-741-32 Sequence 32, Appl
4	269	6.1	540	3	US-08-973-462-22 Sequence 22, Appl
5	267	6.1	688	3	US-09-141-047-8 Sequence 8, Appli
6	250.5	5.7	630	3	US-08-973-462-9 Sequence 9, Appli
7	236	5.4	703	4	US-09-252-991A-17865 Sequence 4, Appli
8	236	5.4	1596	4	US-08-978-277A-4 Sequence 2, Appli
9	235	5.3	1346	2	US-08-635-121-2 Sequence 2, Appli
10	235	5.3	1346	4	US-08-978-277A-2 Sequence 2, Appli
11	235	5.3	2137	4	US-09-134-001C-4463 Sequence 4463, Ap
12	233.5	5.3	2482	4	US-09-252-991A-16967 Sequence 16967, A
13	232.5	5.3	1848	3	US-08-296-791-6 Sequence 6, Appli
14	232.5	5.3	1848	5	PCT-US95-10661A-6 Sequence 6, Appli
15	230.5	5.2	1018	1	US-08-072-610-2 Sequence 2, Appli
16	230.5	5.2	1018	2	US-08-719-822B-2 Sequence 2, Appli
17	230.5	5.2	1018	3	US-09-092-458-2 Sequence 2, Appli
18	217	4.9	1231	3	US-08-904-263A-4 Sequence 4, Appli
19	217	4.9	1231	4	US-09-434-123A-4 Sequence 4, Appli
20	216.5	4.9	1075	4	US-09-252-991A-18387 Sequence 18387, A
21	213.5	4.8	1601	4	US-09-345-473E-40 Sequence 40, Appl
22	213	4.8	1702	3	US-08-296-791-5 Sequence 5, Appli
23	213	4.8	1702	5	PCT-US95-10661A-5 Sequence 5, Appli
24	212.5	4.8	2756	1	US-08-375-709-11 Sequence 11, Appl
25	212.5	4.8	2756	1	US-08-752-929-11 Sequence 11, Appl
26	212.5	4.8	2756	3	US-09-090-793-7 Sequence 7, Appli
27	212.5	4.8	2756	4	US-09-231-899-7 Sequence 7, Appli

28	211.5	4.8	3788	4	US-09-336-447A-76 Sequence 76, Appl
29	210.5	4.8	1140	4	US-07-757-022B-104 Sequence 104, App
30	210.5	4.8	1363	4	US-07-757-022B-52 Sequence 52, Appl
31	210.5	4.8	1404	4	US-07-757-022B-2 Sequence 2, Appli
32	210.5	4.8	1404	4	US-07-757-022B-62 Sequence 62, Appl
33	210	4.8	941	4	US-07-757-022B-14 Sequence 14, Appl
34	210	4.8	1022	4	US-07-757-022B-84 Sequence 84, Appl
35	210	4.8	1038	4	US-07-757-022B-74 Sequence 74, Appl
36	210	4.8	1049	4	US-07-757-022B-58 Sequence 58, Appl
37	210	4.8	1270	4	US-07-757-022B-44 Sequence 44, Appl
38	210	4.8	1311	4	US-07-757-022B-142 Sequence 142, App
39	210	4.8	1313	4	US-07-757-022B-50 Sequence 50, Appl
40	210	4.8	1314	4	US-07-757-022B-46 Sequence 46, Appl
41	210	4.8	1320	4	US-07-757-022B-60 Sequence 60, Appl
42	210	4.8	1320	4	US-07-757-022B-40 Sequence 40, Appl
43	210	4.8	1354	4	US-07-757-022B-48 Sequence 48, Appl
44	210	4.8	1361	4	US-07-757-022B-40 Sequence 40, Appl
45	209.5	4.8	1780	1	US-08-769-309A-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-28749
; Sequence 28749, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28749
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28749

Query Match	9.0%	Score 398.5;	DB 4;	Length 975;
Best Local Similarity	23.1%	Pred. No. 1.6e-20;		
Matches 240;	Conservative 136;	Mismatches 378;	Indels 285;	Gaps 44;
Qy	14	TKFTDC--TRSNRIQPTTHRGYILKNRQIKLIAASVAAVAFQAAH-GLGGLNIQNL	69	
Db	35	TASTDALPLVKREKILKP--RDYTWVRUTLVRAIAAASVLTSMAHGLGLGEITLKSA	92	
Qy	70	DEPFGSITVTGEERAKALLGGGSVTVS---EKGITAKVHKL-----GDKA 111		
Db	93	NQPLDAETELL--EVRO-LGSSEVTPSLASPEFSKAGVDRLYLTLDTKFTPVVKPNCKS	149	
Qy	112	VIATVSSQAVRDPVLVFRIGA---GAQVRYTALIDPVGSPKTKSALSOGKTHRTAP	167	
Db	150	VIRVTSSKVPQEPYINFLVQVLPNGRLIREYTVLLDPLPSQAAASAPQAPV---SAP	206	
Qy	168	TAEQENQAKALRTDKDSANAIVK---PAYNGKTHTVKGETVKQIAAIRPKHLTL	224	
Db	207	RATG-----APRAPQAPVRTTAPAGSDTVTV--SNDLFWIAQRNRTDRVSV	254	
Qy	225	EQVADALLKANPNVSAHG---RLRAGSVLHPLNLRKIAEQPKOTAKPKAETAS-----	276	
Db	255	PQMLAFQELNPGAFVDGNINRLKSGQVLRPTQOMLERSPREALSQVQANQSWRGSR	314	
Qy	277	-----MPSEPSK-QAT-----	286	
Db	315	NPAAGSAGARQIDATQRNAAGSAPFSKVDATNLRVLSGEGKASKGADKGGKDSKALADT	374	

```

287  - - - - -VEKPEKPEAKVA- - - - -299
375  LAVTKESLDSTRRENEELQSRMDLQSLDKLQKLIQLKDAQKLGQLGQAGGQAQF 434
300  - - - - -APAKAEKPAVRPEPVPAAATAAETAESAPOAQAASADITPDE--TGNA 349
435  NAALPDASQFNAAQAQAPQGTPTAAAATTPAPAGEAPAPAQFPVAPPAAEKPPAPA 494
350  VSEPVQVSABEETESGLFGGSYTLLAG- - - - -GGAALITALLLLLRLAQSKRAARTE 402
495  VPAPAPVQAABQAPAPSEF- - - - -DELLANPLWLAVIGGSALLLVLLMLTSRRNAQKEK 549
403  ESVP- - - - -EESPDLDAADDGI-EITFAEVETPATPEBAPKDVNDTTLALDGESEEE 454
550  EEUQAFADAGEEGEDALDKGDPDUTLDEPE-PQVAAAAPQVEKTTAQTSDALGEAD 608
455  L- - - - -SAKOTFDVETTPSNRIDLDFDSLAA-AQNGILSGALTQDEETQKRADAD 504
609  IYIAYGRFNQAAELLQNAIYDEP-QRTDLRLKLMVEYAEMGDREGFARQENELREIGGAQ 667
505  WNALTESTDVVEPTETFPYNVEIVIDPPEPESVAQTAENKEETVDTFDSNLPNNHIG 564
668  -PQVEQLKSRY- - - - -PAMVAVAAVAGLAGKAQDELDSFSLD-DUSLDDSGH-- 714
565  TEETASAKP-ASPSGLAGFLKASSPETILEKTVAEVOQTFE-ELHDFLKVYETDAVAETAP 622
715  - - - - -AAKPDAAAGQULDDAPFLSLDDLGGDDVQADLKSDSGALDDLTLDSLDLLAATA 769
623  ETDPFNAAADDLSALLQPAEAPSVENITEVTAETPDFNATADDLSALLQPSSEVPAVEN 682
770  DKP- - - - -VDLDDPGLFOAE- - - - -LAETPS-OPKHDDLDGDFSLDLDAEP--BDK 810
683  AAETIVADDLSALLO--PAEAPAVEENVTEVAETSDPHTAAD--DLGALLQPAEVPAVE 737
811  LSD-- - - - -DDFLLSUNDEVPAAPADNEFTLDTAAEAPALSUPDDPFLSLADEPTEPAPE 867
738  ENVTKTVAETPDFNATADDLSALLQPSVEVPAVEENAAEITLETPOSNT--SEADALPDF 794
868  KGEDSFAAQLDEVSQAQLDELAS- - - - -NLDPEKSATPSFSAEDAAVAS 910
795  LKQGEETVDWSIYLSEENIPNNADTGFPPSSVSGDAPSEBAKYDILAEMYLEIGRDDAAAE 854
911  ALDGD- - - - -ADDDPDFLSGADEAAT--KLDIARAYIDMGSEGARD 950
855  TVQKLLREAEGDVULKRAQA 873
951  ILDEVL--AEGNDSQQAEEA 967

```

```

RESULT 2
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIS, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

RESULT 3
US-08-714-741-32
: Sequence 32, Application US/08714741
: Patent No. 6500613
: GENERAL INFORMATION:
: APPLICANT: Briles, David E.
: APPLICANT: McDaniel, Larry S.
: APPLICANT: Swiatlo, Edwin
: APPLICANT: Yotter, Janet
: APPLICANT: Crain, Marilyn J.
: APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca
 APPLICANT: Brooks-Walter, Alexis
 TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
 TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
 TITLE OF INVENTION: PORTIONS AND PRODUCTS
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtis, Morris & Safford, P.C.
 STREET: 530 Fifth Avenue
 CITY: New York
 STATE: New York
 COUNTRY: U.S.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patencin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/714,741
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Frommer Esq., William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454312-2460
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8991 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: amino acid
 US-08-714-741-32

Query Match	6.6%;	Score 289;	DB 4;	Length 8991;
Best Local Similarity	21.6%;	Pred. No. 4.1e-11;		
Matches 198;	Conservative 92;	Mismatches 313;	Indels 312;	Gaps 34;
Qy	146	VGYSPKTKSALSOGKTHRKTTAPTAESENQNAKALRKT--DKDSANAAVKPAYNGKTHT	203	
Db	7741	VEXFKT-----DAEQTEQYLAAAEKDLADKKAELEKTEADLLKAVNEPEKPAETPAPA	7795	
Qy	204	VRXGETVKQIAAAIRPK-----HLTLEQVADALLKANPNVSAHGRRLRAGSVLHIFNLN	256	
Db	7796	PKEQPAEQKPPAPAPQAPAPKPEKTDQQAEE-----EDYARRSEEEYNRLFQQQ	7846	
Qy	257	RIKAEQPKP-----QTAKPKAETASMPSEPSKQATVEKPVKEPKEAKVAAP	301	
Db	7847	PPKAEKPAPAPKPEQVPVPAEXPENPAPAPKAPAXAPQPLKPEEPAEQPKP-EKDEEPAQGP	7905	
Qy	302	E-----AKAEKPAVRP---EPVPAANT-----	320	
Db	7906	EPEKPDQQAGEDYARRSGGEYNRFPOQPPKAEKPAPAPKPEQVPAPKPTLLKKAKLAG	7965	
Qy	321	AASETAAESA-----PQAAAAASAIPTPDETGTNAV	350	
Db	7966	AKSAATKKAELPELEKAEAELENLLSTLDPGKQTQDELDKGAERLKNKVEALPNQV	8025	
Qy	351	SEPVEQVSABEETBSGLFGGSYTLLLAGGGAALIALILLRLAQSKRARTEESVPDEEP	410	
Db	8026	SELEEELSKLEDNLKDAETNVEDYIEG-----LEBAIYKQAELEKTPKELDA	8075	
Qy	411	DLDDAADDGHEITFAEVETPA-----TPEPAKNDVN-DTILALDG---ESEEE-454		
Db	8076	ALNELGPDGDEEETPPPEAPAEQPKPEKPAETPAPAPKPEKSADQABEDYARRSEEEY	8135	
Qy	455	--LSAQOTFDVETDTPS-----NRIDLDFDSL-----	479	
Db	8136	NRLTQOOPPPKAEKPAPAPKPEOPAPAPKRSGLIATKKLNLAEARTELLKKKLGLEPGL	8195	

RESULT 4

```

US-08-973-462-22
; Sequence 22, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-22

```

	Query Match	6.1%	Score 269;	DB 3;	Length 540;
	Best Local Similarity	20.9%;	Pred. No. 1.7e-11;		
	Matches 131;	Conservative 112;	Mismatches 262;	Indels 122;	Gaps 19;
Qy	166	APTAES--QBNQAKALRKTDKDSANAAVKPAPYNGKTHTVRKGE-----TVKQIAAAIR	218		
Dd	18	APSVEESVEENVEESVAENVVEESVAENVVEESVAENVVEEIVAPVVEIVAP--	75		
Qy	219	PKHLTLEQ-VADALLKA-NPNVSAGHRLRAGSVLHIPNLNRIKAEQPKPOTAKPKAETAS	276		
Dd	76	---TVEEIIPAVSWVESVAPSVESVE-----ENVEESVAENVVEESVAENVVEESVA	122		
Qy	277	MPSEPSQAQTVEKPVKEPEAKVAAPE-----AKAEKPAVRPEVPFAANTAAETAANSAP	331		

Db 123 ENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVSVESVAPSVESVEENVE 182
Qy 332 QEAASALDTPDETGNVSPVQVSAEETESGLFGSYTLALLAGGALLIALLLLR 391
Db 183 ESVAVNEESVAENVEESVAENVEESVAENVEES 216
Qy 392 LAQSKRAARTBESPEE-EPDLDDAADGIEITFAEVETPATPEPAPKNDVNDTLALDGE 450
Db 217 -----VAENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVSVESVAPVEE 271
Qy 451 SEBELSAKQTPDVTETPSNRIDLDFDSLAAQNGILSGALTQDEETQKRADADWNAIES 510
Db 272 SVEE-----NVESVAEN-----VEESVAENVEESVAENVEESVAEN 308
Qy 511 TDSVVEPETFPYNP-VEIVIDTPEPSVAOTAENK-PETVDTDFSNLPSNNHIGTET 568
Db 309 VEEIVAPTVEEIVAPTVEEIVAPSVVSVESVAPSVESVEENVEESVAENVEESVAENVEES 368
Qy 569 ASAKPSPGLAGFLKASSPETILEKTVAEVQTPPEELHDFLKVVETDAVAETAPETPDFN 628
Db 369 V-----AENVEEIVAPTVEEIVAP-TVEEIVAPSVVSVESVAPSVESVEEN 412
Qy 629 AAADDLSALLOPAEAPSVENITVTVAETPDFNATADDLSALLOPSEVPVAPVEENAAEIVA 688
Db 413 -VEESVAENVEESVAENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVA 466
Qy 689 DDLGALLQAPAPAVEENVETVAETSDPHTAADDLSALLOPAPVAPVEENVTKTVAEIP 748
Db 467 PSVSVESVAPSVESVEENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEIV 520
Qy 749 DFNATADDLSALLOPSEVPVAPVEENAAE 775
Db 521 -----VAPSVESVAPSVESVAE 539

RESULT 5

US-09-141-047-8
; Sequence 8, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; TITLE OF INVENTION: Protein Gene
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141.047A
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 8
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
; OTHER INFORMATION: protein.
US-09-141-047-8

Query Match 6.1%; Score 267; DB 3; Length 688;
Best Local Similarity 23.5%; Pred. No. 3.5e-11;
Matches 161; Conservative 101; Mismatches 242; Indels 180; Gaps 29;
Qy 249 VLHIPNLNRIKAEQPKQTAQKPAETASMPSEPSKQATVKEPKPEAKVA-----APEAK 304
Db 24 IMRILNFGNNSDEKVSNEETKVLE-----SLQPAVDNVDNVPSPSEVSGKEENAEVVK 75
Qy 305 AE--KPAVRPEPAAANTA--ASETAAESAPQAAAS-----AIDTDTDTGNVSPVE 355
Db 76 AEDLPADVSGVHSSEVSGKVKSEETPEVKAEDLQPAVDGSIHSSSVSGKVS 135
Qy 356 QVSAEETESGLFGSYTLALLAGGALLIALLLLRLAQSKRAARTTESVPE-REPLDD 414
Db 136 KTSKEEST-----PEVKAEDLQ 153

Qy 415 AADDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKQTFDVTDTSPNRIDL 474
Db 154 AVDSVSHSSEV-----GEKVSETSK-----EENTP-----EV 182
Qy 475 DFDSLAAQNGILSGALTQDEETQKRADADWNAIESTDSVVEPETFPYNPVEIVIDTPE 534
Db 183 KAEDLQPAVDGSIHSSSVSGKVKSTKE-----ESTPEV-KAEDLQ-----AVDDSV 232
Qy 535 -----PESVAOTA--ENKPETVDTDFSDNLP-----SNNHIGTEETASAKPASPGLAG 581
Db 233 HSSSEVSGKVKSEETPEVKAEDLQPAVDGSHSSSVSGKVKSETSK----- 283
Qy 582 FLKASSPETILEKTVAEVQTPPEELHDFLKVVETDAVAETAPETPDFNAAADDLSALLOPA 641
Db 284 --BESTPEVKAEDLQPAVDSDVE-HSSEVGE--KVSETSKEENTPEVRAED-----LQPA 334
Qy 642 EAPSVEN-----ITVTVAETPDFNATADDLSALLOP-----SEV-----PAVEEN 682
Db 335 VDGSHSSSVSGKVKSEETPEVKAEDLQPAVDSSSTHSSSVSGKVKSETSKES 394
Qy 683 AAEIVADDLSALLOPAEAPAVEEN---VTETVAETSDPHTAADDLSALLOPAEVPAVE-- 737
Db 395 TPEVKAED-----LQPAVDGSHSSSVSGKVKSEETPEVKAEDLQPAVDGSHSS 450
Qy 738 -----ENVTKTVAEIPDFNATADDLSALLOPSEVPVAPVEENAAEITLETDPDNTGEADALP 792
Db 451 SSEVSGKVKSEETPEVKAED-----LQPAVDGSHSSSVSGKVKSETSKESSTP 504
Qy 793 DFLKDGSEETVDSIYLSEENIPNNADTSPFSEVSGVSDAPSEAKYDLAEMYLEIGDRDA 852
Db 505 EVKAEDLQPAVDGSHSSSVSGKVK-----SETSKSEETPEVKAEDLQPAVDGSHSS 559
Qy 853 AETVQKLLLEAEGDVLKRAQALQ 876
Db 560 SEVSGKVKSEETPEVKAERV 583

RESULT 6

US-08-973-462-9
; Sequence 9, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 9
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-9

Query Match 5.7%; Score 250.5; DB 3; Length 630;
Best Local Similarity 24.9%; Pred. No. 4.8e-10;
Matches 123; Conservative 96; Mismatches 215; Indels 59; Gaps 20;
Qy 403 ESVPEEPDLDAAADGIEITFAEVETPATPEPAPKNDVNDTLA--LDGSEBELSAKQT 460
Db 4 ESV-DONKLEAEADIKENILLSNIE-----EPKENIIDLNLLNNGQNSKQESVSEN 55
Qy 461 FVETDTPSNRIDLDFDSLAAQNGILSGALTQDEETQKRADADWNAIESTDSVVEPETF 520
Db 56 VQV-SDELFNELLNSVDVNGVEKNIL-----EESQVNDI-FNSL--VKVQSQQ 104

DB	1265	VTYDSEVMGAGCCQKEBTEVQSLSEBEGMETDVEKEKETKPEQVSEERGE	1316
DB		TELEFAX: 212-765-2519	
DB		TELEX:	
DB		INFORMATION FOR SEQ ID NO: 4:	
DB		SEQUENCE CHARACTERISTICS:	
DB		LENGTH: 1596 amino acids	
DB		TYPE: amino acid	
DB		STRANDEDNESS: single	
DB		TOPOLOGY: linear	
DB		MOLECULE TYPE: peptide	
DB		HYPOTHETICAL: NO	
DB		ANTI-SENSE: NO	
DB		FRAGMENT TYPE: internal	
DB		ORIGINAL SOURCE:	
DB		US-08-978-277A-4	
DB		Query Match	
DB		Best Local Similarity 5.4%; Score 236; DB 4; Length 1596;	
DB		Mismatches 168; Conservative 106; Mismatches 281; Indels 277; Gaps 37;	
QY	258	IKAQPKPOTAKPKAETASMPSEPSKQATVEK-PVEKPE	295
DB		151 IHTESPESADEQKGSSASSPEETTCLEKGPLEAFQGEAEGETTSDEKKREGITP	596
QY	296	AKVAAPAEKAEKPAVRPFVPAANTAASETAAESAPQEAASAIIDTPTDTGNV	351
DB		597 WASFKQWTPKKVRPP-----SESDEK-ELEKVKSATLSSTDSTVSEM	641
QY	352	EPVEQVSAEETESGL-----FGSYTLLAGGGAALIALLLLLLAQSKRARRTEESVPE	407
DB		642 DEVKTVGEEKPEPKRRVDTSVSWEALICV-----SSKKRARKASSODE	688
QY	408	EEPDLD-DAADOGIETTFAE-----VETPATPPAPKNDVNDTLALDGE	450
DB		689 GGPRTTLCGSHRAEASDKAEAGTAVPASTQEQOAGSSSPEA-----GSP	737
QY	451	SE-EELSAKQTFD-VETDTPSNRIDLDFSLAAANGILSGALTQDEETQKRADNNAI	508
DB		738 SEGEVSTWESFKRLVTPRKSKSKLE-----EKAEDSSVEQLSTIEPSREES	790
QY	509	ES-----TDSVVEPETFNPNPVEITVDTPPEPSVAQTAE-----NKPET	548
DB		791 KKFIPGRRKKRAGQEQEATSDSGVEINEDPNVAVPLSEYNAVEREKMAEQNTE	850
QY	549	VDTFDFSNLPSNNHI-----GTEETASAKPASPGLAGFL-----KAS	586
DB		851 LPOLLGAVVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISAVTEPLEHTAGEAM	910
QY	587	SP-ETILEKTVAEVQTP-----EELHDFLKVYETDAVAB--TAPETPD	626
DB		911 PPVEEYTEKDIIEAETFPVLTQTLPEKGDHDMVTSEVDFTSVAVTATETSEALRTEVT	970
QY	627	FNAADDLSALLQPAEAPSVENITET-----VAETPD-----FNATADDLS	668
DB		971 EASGAETDMVSQVLSLTDSPDTTAEATPQVEESGVLDTEEBERQOAILQAVADKVK	1030
QY	669	ALLQPSVEPA-----VEENA-----	683
DB		1031 EESQVATQVQRTGSKALEKVEVEBEDSEVLASEKEKDMKPGVPQEGAGHLAGQ	1087
QY	684	AETVAD-----DLGALLQPAEAPAVENVTETVAETSDPHTA	720
DB		1088 SETGQATPSLEVPETADVHDVATQCVTKLQOLMEQAVAPSSSTLIDSTNGSTPLAD	1147
QY	721	ADLSALLQPAEAPAVENVTKTVAE---IPDNATADDLSALLQPSVEPAVVENAAEIT	777
DB		1148 SDTADGTFQODEITDSQDSKATAAVRSQVTEEEAATAKEEPSTLPNNVPAQEEHGEPPG	1207
QY	778	LET--PDSNTSEADALPDFLKD--GEEEFVDSIYLSENI-----PNN--AD	819
DB		1208 RDVLEPTQOELTAAAVFVLAETVQGEQVDM---LDGEKVKKEEVEFVHSGPNSQKAAD	1264
QY	820	TSPFSVSG-SDAPSEAKYDLAEWYLEIGDRAAAE-----TVQKLLLEAE	864

Db 359 SFKRLVTPRKSKSL-----BEKAGRTLTVVAGCPLRSNRVKNLGFPLRNSPDG 411
 QY 486 ILSGALTQDEETQKRADADWNAIESTSVYEPETFPNPNVPEIIVDTPEPESVACTAEN- 544
 Db 412 GRKGQMGREQA-----TVEDSGPVEINEDEPDVAVVPLSEYD 450
 QY 545 --KPETVDTDFSDNLPS-----NNHI-----GTEETASAKPASPGLAGF 582
 Db 451 AVEREKMEAGNAELPSCWGCWVSEELSKTLVHTVSVAVIDGTRAVTSVEERSFSWISAS 510
 QY 583 L-----KASSP-ETILEKTVAEVQTP-----EELHDFLKVYETDAVAE--TA 621
 Db 511 VTEPLEHTAGEAMPVVEVTEKDIABETPVLTOTLPEGKDAHDMVTSEVDFTEAVTA 570
 QY 622 PETPD-----FNAADDLSALLOPAEAPSVEENITET-----VAETPD----- 659
 Db 571 TETSEALRTEEVTRASGAETDTMVSQVLSQVTSPTTETEEATPVQVEGGVLDTEEBERQ 630
 QY 660 ----FNATADDLSALLOPSEVPA-----VEENA----- 683
 Db 631 TQAILQAVADKVK---BESQVPATQTVORTGSKALEKVEEVEEDSEVLASEKEKDVMPKG 687
 QY 684 -----AEIVAD-----DLSSALLOPAEAPAVEENV 707
 Db 688 PVQEGAGHLAGQSETGQATPESLEVPEVTADVHDVATCQVIKLQQLMEQAVAPESSETL 747
 QY 708 TETVAETSDFHTAADDLSALLOPAEPAVEENVTKTVAE---IPDFNATADDLSALLOPS 764
 Db 748 TDSSTNGSTPLADSDTADGTQDETIDSQSKATAAARQSQVTEEEAATAQKEEPTLNP 807
 QY 765 EVPAVEENAAEITLET--PDSNTSEADALPFLKD--GEETVDSIYLSSENI----- 814
 Db 808 NVPAQEGHEGPGRDVLEPTQOEALAAAAPVWQKTEVQGEVDW---LDGEKVKERQEV 864
 QY 815 -----PNN---ADTSPSESYG--SDAPSEAKYDIAEMYLEIGDRDAAE-----TVQKL 859
 Db 865 FVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSEEGEMETDVEKEKRETKPEQV 924
 QY 860 LEEAE 864
 Db 925 SEEGE 929

RESULT 10

US-08-978-277A-2
 ; Sequence 2, Application US/08978277A
 ; Patent No. 6582956
 ; GENERAL INFORMATION:
 ; APPLICANT: Gelman, Irwin H.
 ; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10112-0228
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,277A
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION NUMBER:
 ; APPLICATION NUMBER: 08/665,401
 ; FILING DATE: 18-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Richard S
 ; REGISTRATION NUMBER: 26,154

; REFERENCE/DOCKET NUMBER: A30558 - 165/34008
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-408-2558
 ; TELEFAX: 212-765-2519
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1346 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEetical: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:
 ; US-08-978-277A-2

Query Match 5.3%; Score 235; DB 4; Length 1346;
 Best Local Similarity 20.2%; Pred. No. 2e-08;
 Matches 171; Conservative 102; Mismatches 268; Indels 304; Gaps 37;

QY 258 IKAEQPKPOTAKPKAETASMESEPSKQATVEK-EVEKPEAKVAAPAEKAEKPAVRPEP-- 314
 Db 151 IHTESPESADEQKGESSASSPEEPETTCLEKGPLEAPM-----GKLRKELLRGEKKR 204
 QY 315 -----VPAANTAASETA---AESAPQ---EAAASAITDPTDETGNVASEPVEQVSADEE 362
 Db 205 KDHSGLIGLOKDGDTQETVRRPSESDEKEELEKVKASATLSSTDSTVSEMQDEVKTVGEQK 264
 QY 363 TE-----SGLFGGSYTLALLAGGGAALFALLLLRLA 393
 Db 265 PEEPKRVRDTSVSEALICVSSKKRARKASSDIRPTLL---GGGQ-----S 310
 QY 394 QSK-----RARRTERSVPEEPFDLDDAADGIEITFAEVETPATPEPA----- 436
 Db 311 QSRGGQQRQSRRTDAVPASTQEQDQA-----QGSSSPPEAGSPSEGEVSTWE 358
 QY 437 -----PKNDVNDTLALDGESEBELSAKQTFDYETDTP--SNRID--LPDLSLAAQNG 485
 Db 359 SFKRLVTPRKSKSL-----EKEAGRTLTVVAGCPLRSNRVKNLGFPLRNSPDG 411
 QY 486 ILSGALTQDEETQKRADADWNAIESTSVYEPETFPNPNVPEIIVDTPEPESVACTAEN- 544
 Db 412 GRKGQMGREQA-----TVEDSGPVEINEDEPDVAVVPLSEYD 450
 QY 545 --KPETVDTDFSDNLPS-----NNHI-----GTEETASAKPASPGLAGF 582
 Db 451 AVEREKMEAGNAELPSCWGCWVSEELSKTLVHTVSVAVIDGTRAVTSVEERSFSWISAS 510
 QY 583 L-----KASSP-ETILEKTVAEVQTP-----EELHDFLKVYETDAVAE--TA 621
 Db 511 VTEPLEHTAGEAMPVVEVTEKDIABETPVLTOTLPEGKDAHDMVTSEVDFTEAVTA 570
 QY 622 PETPD-----FNAADDLSALLOPAEAPSVEENITET-----VAETPD----- 659
 Db 571 TETSEALRTEEVTRASGAETDTMVSQVLSQVTSPTTETEEATPVQVEGGVLDTEEBERQ 630
 QY 660 ----FNATADDLSALLOPSEVPA-----VEENA----- 683
 Db 631 TQAILQAVADKVK---BESQVPATQTVORTGSKALEKVEEVEEDSEVLASEKEKDVMPKG 687
 QY 684 -----AEIVAD-----DLSSALLOPAEAPAVEENV 707
 Db 688 PVQEGAGHLAGQSETGQATPESLEVPEVTADVHDVATCQVIKLQQLMEQAVAPESSETL 747
 QY 708 TETVAETSDFHTAADDLSALLOPAEPAVEENVTKTVAE---IPDFNATADDLSALLOPS 764
 Db 748 TDSSTNGSTPLADSDTADGTQDETIDSQSKATAAARQSQVTEEEAATAQKEEPTLNP 807
 QY 765 EVPAVEENAAEITLET--PDSNTSEADALPFLKD--GEETVDSIYLSSENI----- 814
 Db 808 NVPAQEGHEGPGRDVLEPTQOEALAAAAPVWQKTEVQGEVDW---LDGEKVKERQEV 864

QY 691 LSALLOPAEAPAEVENVETVAETSDPHTAADDLSALLOPAEAPAEVENVETKTVAEIPDF 750
Db 1454 ----ROP-EANTVADNSVANNSESE--SKSRRRSVQPKETSABETTVASTQ-----1500
QY 751 NATADDLSALLOPSE---VPAVEENAAEITLETDPDSNTSEADALPDFLKDGEETVDWSI 807
Db 1501 ETTVDNSVSTPKPRSRTRRSVQNTSYE-FVELPTENAENAVNQ-----1544
QY 808 YLSEENIPNNAD-----TSPFESVSGDAPSEAKYDLAEMYLEIGDRDAAAEVTKVQLLEE 862
Db 1545 --SGNVANSOPALRNLTSKNTNAVISNAMAQAF-----VALNVG--KAVSQHISOLEMN 1596
QY 863 AEG 865
Db 1597 NEG 1599

RESULT 14
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleht, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-6

Query Match 5.3%; Score 232.5; DB 5; Length 1848;
Best Local Similarity 19.2%; Pred. No. 5e-08;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;
QY 35 LKNRQIKLIAASVAAASFOAHAG-----LGGINI-QSNLDEPFSGSIIVT 80
Db 766 LYSGRNVANITSNITASNAAQVHGYKGTGTVCVRSYDTGYVTCNLSNLSKALNSFNPT 825
QY 81 GEEAKA-LILGGSVTVSSEKGLTAKVHKLGDRAVIAVSEQAQRDPVLVFRICAGAQVREY 139
Db 826 NLRGNVLTENASFTLGRANLEFGTIGTIGTQVNLKNSH-----WHLTGSNVNQL 877
QY 140 TAILDPVGYSPKTSALS DGKTHRKTAFTAESQENQNAKLRK-----TDKK 186

Db 878 N-----LTNGHIIHLNAQNDANKVTTNTLTVNSLNGSGSFYVWDFNNK 922
QY 187 DSANAAVKPAYNGKTHV--RKGETVKQIAAAIRPKH--LTLRQVADA-----LKKANP 236
Db 923 SNKVVNKGATGNFTLQVADKTGE-----PHNELTLFDASNATRNMLEVTLANG 972
QY 237 NV-----SAHGR-----RAGSVLHPNLRKAKBQPKOTAK---PKA 272
Db 973 SVDRGAWKYKLRVNGRYDLYNPEVEKRNQTVDTNITTPNDIQADAPSACSNNNEIARV 1032
QY 273 ETASMSSEPSKQATV--EKDVEKPEAKVAPEAKAEKPAVRPEVPAAATAA-----322
Db 1033 ETPVPPAPATESAIASEQETPR-AETAQPAEMEETNTANSTETAPKSDTATOTENPSE 1091
QY 323 ---SET-----AESAPOEAAAASAIIDTPTDGTGNVSEPVQVS 358
Db 1092 SVFSETTEKVAENPPQENETVAKNEQATEPTFONGEVAKEDQPTVEANTQTNEATQSEG 1151
QY 359 ABEETESGLFGGSYTLALLAGGAALIALLLRLAQSRRARTEESVPEEPD--LDDAA 416
Db 1152 KTEETQT-----AETKSEPTESVTVSENQPEKTVSQST 1184
QY 417 DDGIEI---TFAEVET-----PATPEAPKNDVNDTLALDGESEBELSAKQ 459
Db 1185 EDKVVWEKEKAKVETEETQKAPQVTSKEPPKQAEPAPEEVPTDTNAAEEAQAQQOFTT 1244
QY 460 TFDVETDTPSNRIDLDLDFSLAAQNG-----ILSGALTODEETOKRADADWNAIBSTDS 513
Db 1245 VAAAEITSPNSKPAEETQOQSEKTNAPVTPVVSSENTATQPTETEETAKVE---KEKTQE 1301
QY 514 VYBPETENPNVPEIVIDTPEPESVAQTAENKPEVTDTDFSDNLPNNHIGTEETAQAKP 573
Db 1302 V--PQVASQESPKQ---EQPAKPEQAT---KPOAEP--RENVLTITKNVG-EPQPOAQP 1350
QY 574 ASPS---GLAGFLKASSPETILEKTVAEVQTPBELHDFLKYETDAVAETAPETDFNAA 630
Db 1351 QTOSTAVPTTGTETAANSKPAKQQAQKQTEPAENVTNTKEPOSQTSATVSTQPA 1410
QY 631 ADDLSALLOPAEAPSVENITETVAETPDNFNATADDLSALLOPSEVPAVEENAAEIVADD 690
Db 1411 KETSSNVEQAPENSINTGSATMTET-----AEKSDKQME---TVTEND 1453
QY 691 LSALLOPAEAPAEVENVETVAETSDPHTAADDLSALLOPAEAPAEVENVETKTVAEIPDF 750
Db 1454 ---RQ- EANTVADNSVANNSESE--SKSRRRSVQPKETSABETTVASTQ-----1500
QY 751 NATADDLSALLOPSE---VPAVEENAAEITLETDPDSNTSEADALPDFLKDGEETVDWSI 807
Db 1501 ETTVDNSVSTPKPRSRTRRSVQNTSYE-FVELPTENAENAVNQ-----1544
QY 808 YLSEENIPNNAD-----TSPFESVSGDAPSEAKYDLAEMYLEIGDRDAAAEVTKVQLLEE 862
Db 1545 --SGNVANSOPALRNLTSKNTNAVISNAMAQAF-----VALNVG--KAVSQHISOLEMN 1596
QY 863 AEG 865
Db 1597 NEG 1599

RESULT 15
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York

STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,610
FILING DATE: 19930602
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 5986/07686
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
C-terminal
FRAGMENT TYPE: C-terminal
ORGANISM: Plasmodium vivax
IMMEDIATE SOURCE:
CLONE: pMB3.3.1

US-08-072-610-2

Query Match 5.2%; Score 230.5; DB 1; Length 1018;
Best Local Similarity 21.4%; Pred. No. 2.9e-08;
Matches 163; Conservative 97; Mismatches 301; Indels 201; Gaps 32;
QY 158 DGKTHRTAPTAEQENQKALRKTKKDSANAAVPAYNGKTHTVRKGETVVKQIAAAI 217
Db 192 DAETEGATGDAETENG-----ATVYVDTEDSSADGAERKVVPAQENQVADNSDALFGSI 247
QY 218 RPHELTLEQVAD--ALLKANPNVSAHGLRAGSVLHPLNLRKAEQPKQTAQPKAETA 275
Db 248 LDKDIIIFDHKDPFLPE-----QIVAGTAKHTVG-----Q 278
QY 276 SMPSEPSQATVEKPKPAKVAAPKAPKPAVRPEPVPAANTAAETAESAPOEAA 335
Db 279 ELPKMPVPLVAEPFAQVP-----AEELDATPEDDFALDVTSPEERVELVLDEEA 328
QY 336 ASADITDTETGNVSPVEQVSAEETESGLFGSVTLILAGGGAALIALLLRLAQS 395
Db 329 TEEST---EVGPTTEGTTELDATPEDG-----FRIRN 360
QY 396 KRAARTEESPEEPDLDDAADDGI--EITPAEVEPTATPEPAKNDVNDTLALDG--- 449
Db 361 CR-RRNRNVEGEE--TEEAAGVSEETPEGELEATPE-----DDFALDGTLE 409
QY 450 ESEELSAKQTFDVTETPSNRIDLDFDSLAAQAQNGILSGALQTDE-----ET 497
Db 410 ETEETAEGEETVEGE-ETVEGEETVEGEAAEAGEEL--EATPEDDFQLEEPSGEGEG 466
QY 498 QKRADADWNALESTDSVPEP-ETPNPNPVIIVDTPEPSVAQTAENKPKETVDTDFSDN 556
Db 467 EGEGEGEALVAVPVVAEPVEVVTPAQPVK-----PMVAPTAD---ETLFDVILDN 515
QY 557 LPSNNHIGTETASAKPASPSGLAGFLKASGPETILEKTV-----EVQTPPELHDLKV 611
Db 516 DLTVADITSPEPLFKQILKPDAGEAVTVPSKEAPVQVPVAVGAQVEPTEE---LMQL 571
QY 612 YETDAVAETAPETP-----DFNAAADDLSALIQAAPASVE 647

Search completed: December 12, 2003, 17:33:40
Job time : 26 secs

Db 572 QEDDFELGTAAPAEPEGELVLEGEGETEERPREGEPTGEVPEELBA--TPEDDFELE 629
QY 648 ENITETVAETPDPFNATADDLSALLOPSEVPAVENAAEIVADDLSALLOPAEAAVENV 707
Db 630 EPTGERVEETVEGETAEGEEVEEVPAAVEEVEEVPABV-----EEVEEVPPEV 678
QY 708 TETVAETSDFHATAADDLSALLQPAEVPVENVTVAEIPDFNATADDLSALLQPS--- 764
Db 679 EEVPAKVEEVEEVEEVE------EVPEEVEEVEEVEEVEEVEEVEEVPAAVE 732
QY 765 -EVPAAVENAAEITLETPODNTSADALPDLKDGEEETVDWSTYLSSEENIPNNADTSFP 823
Db 733 VEVPAAVVEE-----EVPEEVEEVEE-----EPEPVE-----EEDV---LQLVIP 768
QY 824 S-ESVGSADAPSEAKYDLAEMVLEIGDRDAAAETVQKLLLEAE 864
Db 769 SEEDIQLDKPK--KDELGSGILSIIDMH-YQDVPKPEMEEE 807

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:32:32 ; Search time 37 Seconds
(without alignments)

4423.395 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCFPMWTKFTDCT.....EEAEGDVLRQAALQELGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	358	8.1	1095	15 US-10-128-714-8305	Sequence 8305, Ap
2	315.5	7.2	1786	10 US-09-742-096-3	Sequence 3, Appli
3	315	7.2	1000	15 US-10-128-714-3305	Sequence 3305, Ap
4	269	6.1	540	10 US-09-742-096-22	Sequence 22, Appl
5	253	5.7	1367	10 US-09-801-368-108	Sequence 108, Appl
6	250.5	5.7	630	10 US-09-742-096-5	Sequence 5, Appli
7	244	5.4	1616	11 US-09-820-843A-16	Sequence 16, Appl
8	238	5.4	2665	9 US-09-864-761-34248	Sequence 34248, A
9	236	5.4	3664	15 US-10-177-293-423	Sequence 423, Appl
10	236	5.4	1596	10 US-09-902-432-4	Sequence 4, Appli
11	236	5.4	2478	9 US-09-815-242-5816	Sequence 5816, Ap
12	236	5.4	2478	9 US-09-902-432-2	Sequence 2, Appli
13	235	5.3	1346	10 US-09-815-242-12967	Sequence 12967, A
14	232.5	5.3	1848	11 US-09-839-996-6	Sequence 6, Appli
15	232.5	5.3	1848	15 US-10-080-505-6	Sequence 6, Appli

Sequence 5635, Ap
Sequence 12389, A
Sequence 5644, Ap
Sequence 8240, Ap
Sequence 70, Appl
Sequence 253, Appl
Sequence 11938, A
Sequence 33686, A
Sequence 3, Appli
Sequence 13, Appl
Sequence 10395, A
Sequence 909, Appl
Sequence 40, Appl
Sequence 5, Appli
Sequence 5, Appli
Sequence 7, Appli
Sequence 76, Appl
Sequence 4, Appli
Sequence 104, Appl
Sequence 52, Appl
Sequence 2, Appli
Sequence 62, Appl
Sequence 14, Appl
Sequence 84, Appl
Sequence 74, Appl
Sequence 58, Appl
Sequence 44, Appl
Sequence 42, Appl
Sequence 142, Appl
Sequence 50, Appl

228.5 5.2 2368 9 US-09-815-242-5635
228.5 5.2 2368 9 US-09-815-242-12389
227.5 5.2 2472 9 US-09-815-242-5064
226.5 5.1 1178 15 US-10-128-714-8240
226.5 5.1 1879 10 US-09-971-536-70
226 5.1 1539 12 US-10-085-959-253
225.5 5.1 582 15 US-10-156-761-11938
225 5.0 980 12 US-10-029-386-33686
221 5.0 1781 9 US-09-738-877-3
219 5.0 1781 11 US-09-961-403-13
219 5.0 1258 15 US-10-156-761-10395
216 4.9 1521 12 US-10-238-075-909
214 4.9 1601 10 US-09-862-027-40
213.5 4.8 1702 11 US-09-839-996-5
213 4.8 1702 15 US-10-080-505-5
213 4.8 2756 15 US-10-331-061-7
212.5 4.8 3788 11 US-09-952-267-76
211.5 4.8 2283 12 US-10-172-502-4
211 4.8 1140 14 US-10-124-557-104
210.5 4.8 1363 14 US-10-124-557-52
210.5 4.8 1404 14 US-10-124-557-2
210.5 4.8 1404 14 US-10-124-557-62
210 4.8 941 14 US-10-124-557-14
210 4.8 1022 14 US-10-124-557-84
210 4.8 1038 14 US-10-124-557-74
210 4.8 1049 14 US-10-124-557-58
210 4.8 1270 14 US-10-124-557-44
210 4.8 1311 14 US-10-124-557-42
210 4.8 1313 14 US-10-124-557-142
210 4.8 1314 14 US-10-124-557-50

ALIGNMENTS

RESULT 1

US-10-128-714-8305
; Sequence 8305, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengji
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8305
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1095)..(1095)
; OTHER INFORMATION: X= any amino acid
US-10-128-714-8305

```
Query Match      8.1%; Score 358; DB 15; Length 1095;
Best Local Similarity 22.1%; Pred. No. 5.5e-14;
Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;

Db      1056 QVPAQDEADVAAIEEQEAEVAPVTKPAAPAAE 1090

RESULT 2
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773US0DIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

Query Match      7.2%; Score 315.5; DB 10; Length 1786;
Best Local Similarity 20.2%; Pred. No. 5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY      64 NISNLDPEPSGSIIVTGGEEAKALLGGSVTVSEKGLTAKVHKLGDKAV---IAVSSEQA 120
DB      242 NVENVEENDGGSVASSVEES-----IASSVDESIDSIEENVAPTVEEI 286
QY      121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTSALSDGKTHRKTAFTAES--OENONAK 178
DB      287 VAPSVV-----ESVAPSVESVEENVEES 310
QY      179 ALRKTDDKDSANAARKPAYNGKTHVRKGE-----TVKQIAAAIRPKHLLEQ-VADALL 232
DB      311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY      233 KA-NPNVSAHGRLAGSVLHPNLRILKAEQPKQTAKPAAETASMPSESKQATVSKPV 291
DB      365 ESVAPSVESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY      292 EKPEAKVAAP-----AKAEKPAVRPFPVPAANTAASETAAESAPQAAAAAIDPTDET 346
DB      416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVESVEENVEESVAENVEESVAENV 475
QY      347 GNAVSEPVQVSAEETESGLFGSSYTLLLAGGGAALIALLLLRLLAQSKRARTESVP 406
DB      476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY      407 EE-EPDLDADAOGIEITAEVETPATPEPAPKNDVNDTLALDGESEELSASAKOTFFDET 465
DB      505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVESVEE-----NVEE 557
QY      466 DTFSNRIDLDFDSLAAQNGILSGALTQDBETQKRADADWNAIESTDSVYEPETFNPNP 525
DB      558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY      526 -VEIVIDTPPESVAQTAENK-PETVDTDFSDNLPNNHIGTETTAQKAPSPGLAGFL 583
DB      602 TVESIVAPSVVESVAPSVVESVEENVEESVAENVEESVAENVEESV-----647
QY      584 KASSPETILEKTVAEQTPEELHDFLKVYETDAVETAPETPDFAAADDLSALLQAPAA 643
DB      648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVVESVEEN-VEESVAENVEESVA 704

Query Match      7.2%; Score 315.5; DB 10; Length 1786;
Best Local Similarity 20.2%; Pred. No. 5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY      64 NISNLDPEPSGSIIVTGGEEAKALLGGSVTVSEKGLTAKVHKLGDKAV---IAVSSEQA 120
DB      242 NVENVEENDGGSVASSVEES-----IASSVDESIDSIEENVAPTVEEI 286
QY      121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTSALSDGKTHRKTAFTAES--OENONAK 178
DB      287 VAPSVV-----ESVAPSVESVEENVEES 310
QY      179 ALRKTDDKDSANAARKPAYNGKTHVRKGE-----TVKQIAAAIRPKHLLEQ-VADALL 232
DB      311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY      233 KA-NPNVSAHGRLAGSVLHPNLRILKAEQPKQTAKPAAETASMPSESKQATVSKPV 291
DB      365 ESVAPSVESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY      292 EKPEAKVAAP-----AKAEKPAVRPFPVPAANTAASETAAESAPQAAAAAIDPTDET 346
DB      416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVESVEENVEESVAENVEESVAENV 475
QY      347 GNAVSEPVQVSAEETESGLFGSSYTLLLAGGGAALIALLLLRLLAQSKRARTESVP 406
DB      476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY      407 EE-EPDLDADAOGIEITAEVETPATPEPAPKNDVNDTLALDGESEELSASAKOTFFDET 465
DB      505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVESVEE-----NVEE 557
QY      466 DTFSNRIDLDFDSLAAQNGILSGALTQDBETQKRADADWNAIESTDSVYEPETFNPNP 525
DB      558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY      526 -VEIVIDTPPESVAQTAENK-PETVDTDFSDNLPNNHIGTETTAQKAPSPGLAGFL 583
DB      602 TVESIVAPSVVESVAPSVVESVEENVEESVAENVEESVAENVEESV-----647
QY      584 KASSPETILEKTVAEQTPEELHDFLKVYETDAVETAPETPDFAAADDLSALLQAPAA 643
DB      648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVVESVEEN-VEESVAENVEESVA 704
```


Qy	219	PKHITLEQ- VADALLKA- NPNVSAHRLRAGSVLHI PNLNRIKABQPPQAKPAERTAS	276
Db	76	-----TVEEIVAPSVVSVAPSVE-----ENVEESVAENVEESVAENVEESVA	122
Qy	277	MPSEPSQOATVEKDVEXPEAKVAPE-----AKAEKPAVRPEPVPAAANTAASETAESAP	331
Db	123	ENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVSVAPSVEESVAENVEE	182
Qy	332	QEAASAIIDPTDGTGNVSPVEQVSAEEETESGLFGGSYTLILLAGGAALLIALLLILR	391
Db	183	ESVAENVEESVAENVEESVAENVEESVAENVEESVAENVEESVAENVEES	216
Qy	392	LAOSKRAARTESVPEE-EPDLDDAADGIEITFAEVTPTATPEPAPKNDVNDTIALDGE	450
Db	217	-----VAENVESVAENVEESVAENVEEIVAPTVEEIVAPSVVSVAPSVEE	271
Qy	451	SEBELSAQOTPDVETDTPSNRIDLDFDSLAAQNGILSGALTODRETOKRADADNNAIES	510
Db	272	SVEE-----NVSESVAEN-----VEESVAENVEESVAENVEESVAEN	308
Qy	511	TDSVVEPETFPNYP-VEIVIDTPPEPSVAQTAENK-PETVDTDFSDMLPSNNH:GTBEET	568
Db	309	VEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESVAENVEESVAENVEES	368
Qy	569	ASAKPASPSGLAGFIKASPTILEKTVAEVQTPPEELHDFLKVYETDVAEVAETPTDPDN	628
Db	369	V-----AENVEEIVAPTVEEIVAP-TVEEIVAPSVVSVAPSVEESVAEN	412
Qy	629	AAADLSALLOPAEAPSVVEENITETVAETPDENATADOLLSALLOQSEVPAPVENAAEIVA	688
Db	413	-VEESVAENVEESVAENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVA	466
Qy	689	DDLGSALLLOPAEAPAEVENNETVAETSDFHTAAADLSALLOPAEVPAPVENYTKTVAEIP	748
Db	467	PSVVESVAPSVVEESVAENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI-	520
Qy	749	DFNATADLSALLOQSEVPAPVEENAAE	775
Db	521	-----VAPSVVESVAPSVVESVAE	539

RESULT 5

```

US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
;
; TITLE OF INVENTION: Methods for Improving
; FILE REFERENCE: 109272.147
;
; CURRENT APPLICATION NUMBER: US/09/801,368
;
; CURRENT FILING DATE: 2001-03-07
;
; PRIOR APPLICATION NUMBER: US 09/487,558
;
; PRIOR FILING DATE: 2000-01-19
;
; PRIOR APPLICATION NUMBER: US 60/160,587
;
; PRIOR FILING DATE: 1999-10-20
;
; NUMBER OF SEQ ID NOS: 440
;
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 108
;
; LENGTH: 1367
;
; TYPE: PRT

```

; ORGANISM: *Saccharomyces cerevisiae*
US-09-801-368-108

Query Match 5.7%; Score 253; DB 10; Length 1367;
Best Local Similarity 18.2%; Pred. No. 2.9e-07;
Matches 132; Conservative 130; Mismatches 398; Indels 64

[illegible]

RESULT 6

US-09-742-096-5
Sequence 5, Application US/09742096
Patent No. US2002015441A1
GENERAL INFORMATION:
APPLICANT: DRUILHE, PIERRE
APPLICANT: DAUBERSTES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-EM
FILE REFERENCE: 200773US01V
CURRENT APPLICATION NUMBER: US/09/
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 08/97/
PRIOR FILING DATE: 1998-02-06

GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US 09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 34248
 LENGTH: 2665
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AL034555.2
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.3
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7
 OTHER INFORMATION: EXPRESSED IN HELA100, SIGNAL = 7.7
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
 OTHER INFORMATION: EST HUMAN HIT: AU117052.1, EVALUATE 0.00e+00
 OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUATE 3.00e-10
 US-09-864-761-34248

Db 952 ANSTTDSIQEPVVLFFH-----SRFELTRMOCKEKEKQKQKQEDTTEHPKTPSAP 1007
 Qy 171 SOENQAKALRKTDKK-----DSANAAYKPAVNGKT-----HTVRKGETVKQIA 214
 Db 1008 ENKDSLEKTPSPVGPSPVTVVTVLESAPSALEKTTGDKTVEAPLVEKTEKTEPATVSEAK 1067
 Qy 215 AAIIRPKHLTLEQVADALLK--ANPNVSA--HGRLRAGSVLHPIINLRIRKAEQPKPTAK 269
 Db 1068 PASEPAPAPVEQLEQVLDLPGADPDKEAAMPACVEGSG-----SGDQPPYLDK 1117
 Qy 270 PKAETAS-----MPSEPSKQATVEKPKPEAKVAAPAKAKPAKPAVREPVPVPAATAA 322
 Db 1118 PPTPGASFQAESNVDPDSTQPLSKPAQKSE---EANEKPAKPKPDATADAEPDANOKA 1174
 Qy 323 SETAAESAPQEAASADTPT--DETGN-----AVSEPVQVSAEETE 364
 Db 1175 -EAAPESQPPASELVDPPVAAKDKKPNKSKSKTQVQAAAVSIVKPKVTRKERIDRE 1233
 Qy 365 SGLFGGSYTLLLAGGGAALIALILLRLAQSKRRARTEESVPEEPDLDAAADGIIITF 424
 Db 1234 -----KLKRSNSPRGEAQKLELKMEAEKITRTASKNSA 1267
 Qy 425 AEVETPATBP-----APKNDVNDTLALDGESEEEELSAKQTFDVTDTFSNRIDLDPSLAA 481
 Db 1268 ADLEHPEPSLPLSRTRRRNVSVYATMGDHNRSFVKEPVE-QPRVTRKRLERELQEA 1326
 Qy 482 AQNGILSGALTQDEETOKRADWNALJESTDSVVEPETFP----- 522
 Db 1327 VPTTFRRG---RPPKTRRADEE---EENEKPEATLKPPEGWRSPRSOKTAAGGPG 1379
 Qy 523 --YNPVEIVIDTPPEVS-----AQTAEKNKE 547
 Db 1380 GKKGNEPKVDATPEATTVEGPOIGVKESSEMPKAAEEERAGSEQKDRDKDAGTDKNPPE 1439
 Qy 548 TVDTDFSDNLP-----NNHIGTEETASAK-----PASPGLAGFLKASSPET 590
 Db 1440 TAPVEVVEKKPAPEKNSKRSRSLAVDKSASLKNVDAAVSPRGAA---AQAGER 1495
 Qy 591 ILEKTVAEVQTPPELHDFLKYVETDAVETAPETPDFAAADDLSALLQPAEAPSVE--- 647
 Db 1496 --ESGVAV--SPEKSESFOKEDGLSSQLKSDVPDPKEPEKEDVSASGPSPEATQAKOM 1552
 Qy 648 --ENITETVARTPDNFATA---DLSALLQPS--EVPVAEENAAEIVA-----DLSAL 694
 Db 1553 ELEQAVEHIAKLAESASAAKADAPGLAPEDRDKPAHQASSETELAAAGSIINDISG- 1611
 Qy 695 LQPAEAPAVEENVTTETVAETSDFHTAADDLSALLQPAE-----VPAVEENVTKTVAE 746
 Db 1612 -EPENFPAPPYPGES---QTDLOPPAG--AQALQPSSEGMETDEAVSGILETEAATES 1665
 Qy 747 IPDENATADDLSALLQPSVEPVAEENAAEITLETSDNTSSEADALPDFLKD--GERETV-- 803
 Db 1666 RPPVN--APDPSA--GPTDTTKEARGNSSETSHSVPEAKGSEVEVTLVRKDKGROKTRTS 1721
 Qy 804 -----DWSITLSEENIP--NNADTSPSPSVGS-----DAPSEAK 836
 Db 1722 RKKNTNKKVAVPVESHVPSNQAGSPAANEGTTVQHPEAPOEEK 1768
 RESULT 9
 US-10-177-293-423
 ; Sequence 423, Application US/10177293
 ; Publication No. US20030124128A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lillic, James
 ; APPLICANT: Glatt, Karen
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Gannavarpu, Manjula
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Mertens, Maureen
 ; APPLICANT: Myer, Vic
 ; APPLICANT: Wang, Youzhen

Query Match 5.4%; Score 238; DB 9; Length 2665;
 Best Local Similarity 20.5%; Pred.No. 6.4e-06;
 Matches 182; Conservative 119; Mismatches 352; Indels 234; Gaps 38;
 114 AVSSEQAVRDPVLVFRIGAGQVREYTAILD---PVGYSPKTSALSDBGKTRKTAPTAE 170

APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 423
LENGTH: 3664
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-423

Query Match 5.4%; Score 238; DB 15; Length 3664;
Best Local Similarity 20.5%; Pred. No. 1e-05;
Matches 182; Conservative 119; Mismatches 352; Indels 234; Gaps 38;
QY 114 AVSSEQAVRQVPLVFRICAGQVREYTAIDL---PVGSPKTKSALSDGKTHRTAPTAE 170
DB 1569 ANSTDSIQBPVVLH---SRFELTRMOQKEKQKPEVEKQEDTENHPKTPESAP 1624
QY 171 SQENQAKALRKTDKK-----DSANAAVPAFYNGKT-----HTVKGTVKQIA 214
DB 1625 ENKDSLKTPEPSVGPSPVTVVLESAPSALKTTGDKTVEAPLVEKTVPEPATVSEAK 1684
QY 215 AATPKHLTLQVADALLK--ANPNVSA---HGRLRAGSVLHPINLNRIKAEQPKPOTAK 269
DB 1685 PASEPAPAPVEQLQVLDPLPGADPDKEAAMPAGVEEGS-----SGDQPPYLDK 1734
QY 270 PKAETAS-----MPSEPSQATVEKPEKPAKVAPEAKAEKPAVRPEPVPAAATAA 322
DB 1735 PPTGASFSQASNVDEPDSGTQLSPKQKSE---EANEKPAEKPDATADAEPDANKA 1791
QY 323 SETAASAPQEAASAIIDTPT---DETGN-----AVSEPVQVSAEBETE 364
DB 1792 -EAAPEQPPASEDLEVDPPVAADKKPKNSKRKTVPQAAAVSIVEKPVTRKSERIDRE 1850
QY 365 SGLFGGSYTLLAGGGAALIALLLRLAQSKRARTEESVPEEPDLDDAADDGIBITF 424
DB 1851 -----KLKRNSPRGEAQKLELKMEAEKITRTASKNSA 1884
QY 425 AEVETPATPEP--APKNDVNDTLALDGESEELSAKQTFDEVTDTSTNRIIDLDFSLAA 481
DB 1885 ADLHPSPSLSTRNRNRSVTVATGWDHNRSPVKEPVE--QPRVTRKRLERELQEA 1943
QY 482 AQNGILSGALTQDEBETOKRADWNALSTDSVYEPETENP----- 522
DB 1944 VPTTPRGG--RPKPTRRADEE---EENEKPEAPETLKPPEGWRSRPSOKTAAGGPGQ 1996
QY 523 --YNPFEIVIDTPPEPV-----AQTAEKPE 547

DB 1997 GKKGNEPKVDATRPEATTEVGPQIVKKESSMEPKAAEEBAGSEQKRDKGDTKPNPE 2056
QY 548 TVDITDFSDNLPS-----NNHIGTEETASAK-----PASPSGLAGELKASSPET 590
DB 2057 TAPVEVVEKKPAPEKNSKRSRNSFLAYDKGASLXNVDAAVSPRGAA-----AQAGER 2112
QY 591 ILEKTVAEVQTPPELHDFLKVYETDAVATAPETPDFNAADLLSALLQPAEAPSVB--- 647
DB 2113 --ESGVAV--SPEKSESPOKEDGLSSQLKSPDVPDKEPEKEDVSAGSPSEATOLAKQM 2169
QY 648 --ENITVETPTDFNATA---DDLSSALLQPS--EVPAAVEENAAEIVA-----DDLAL 694
DB 2170 ELEQAVEHIAKLAESASAAKADAPGLAPEDRDKPAHQASSETELAAIGSIINDISG- 2228
QY 695 LOPAEAPAVEENVTVETVAETSDFTTAADDLSALLQPAE-----VPAVEENVTKTVAE 746
DB 2229 -EPENFPAPPYPGES---QTDLOFPAG--AAALQPSEEGMETDEAVSGILETEAATSS 2282
QY 747 IPDNATADLLSALLQPSVEPAAVEENAAEITLETTPDSNTSEADALPDFLKD--GEETV-- 803
DB 2283 RPPVN--APDESA--GPTDTKEARNSSETSHSVPEAKGSEVEVTLVRKDKGKQKTR 2338
QY 804 -----DWSIYLSEENIP--NNADTSPPSVSGS-----DAPSEAK 836
DB 2339 RRKRNKVKVAVPVESHVPSNQAGSPAAANEGETTVQHPPEAPOBEK 2385
RESULT 10
US-09-902-432-4
Sequence 4, Application US/09902432
Patent No. US20020160002A1
GENERAL INFORMATION:
APPLICANT: Irwin H. Gelman
APPLICANT: Susan G. Jaken
TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
FILE REFERENCE: A30558-A-FWC-A 070156.0597
CURRENT APPLICATION NUMBER: US/09/902,432
CURRENT FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: 08/978,277
PRIOR FILING DATE: 1997-11-25
PRIOR APPLICATION NUMBER: 08/665,401
PRIOR FILING DATE: 1996-06-18
PRIOR APPLICATION NUMBER: 08/635,121
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1596
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-902-432-4
Query Match 5.4%; Score 236; DB 10; Length 1596;
Best Local Similarity 20.2%; Pred. No. 4.2e-06;
Matches 168; Conservative 106; Mismatches 281; Indels 277; Gaps 37;
QY 258 IKAEQPKPQAKPAETASMPSPSKQATVEK--PVEKPE----- 295
DB 537 IHTESPESADEQKGSASSPSEPEETTCLEKPLEAPQDGEAEEGTSDGKKREGITP 596
QY 296 ---AKVAAPEAKAEKPAVRPEPVPAAATAASETAESAAPQEAASAIIDTDTGNAVS 351
DB 597 WASFKMVTFKVRRP-----SESDKEE-ELEKVKSATLSTOSTVSEMQ 641
QY 352 EPVEQVSAEBETESGL----FGGSYTLLAGGGAALIALLLRLAQSKRARTEESVPE 407
DB 642 DEVKTVGEEKPPEPKRVRDTSVSWREALICVG-----SSKKARAKASSDDDE 688
QY 408 -----EEDPLD--DAADDGIEITFAE---VETPATPEAPKNDVNDTLALDGE 450
DB 689 GGPRTLGGDSHRAEASKDKAGTDAVPASTQODQAQSSSPEPA-----GSP 737
QY 451 SE-BELSAKQTFD-VETDTPSNRIDLDFDSLAAQNGILSGALTQDEBETOKRADWNAL 508

```
Db 738 SEGEVSTWESFKRLVTPRKSKSLE-----EKAEDSVBQLSTELPSREBS--WVSI 730
Qy 509 ES-----TSVYEPETFPNPNVBEIVIDTPESVAQTAE-----NKDET----- 548
Db 791 KKEIPGRKRKADCKQEQATVEDSGPVEINEDDENPVAVVPLSEYNAVERKERKEAQNTE 850
Qy 549 -----VDTDFSNLPSNNH-----GTEETASAKFASPSGLAGFL-----KAS 586
Db 851 LPOLLGAVYSEELSKVLNHTVSAVIDGTRAVTSVEERSPSWISASVTPLEHTAGEAM 910
Qy 587 SP-ETILEKTVAEQTP-----EELHDFLKVYETDAVE--TAPETPD----- 626
Db 911 PPVEEVTEKDI IAEETVLTQTLPEGKAHDDMVTSVDFTSEAVTATETSEALRTEVT 970
Qy 627 -----FNAADLGSALLOPAEAPSVENITET-----VAETPD-----FNATADDLS 668
Db 971 EASGAETTDWVAVSQLTSDPTTETATPVQEVESGLDTEEBEERTQAILQAVADKVK 1030
Qy 669 ALLQPSVPA-----VEENA----- 683
Db 1031 -----EESQVPAOTVORTGSKALEKVEVEEDSEVLASEKEKDVMPKGPVQEAHQAOG 1087
Qy 684 -----AEIVAD-----DLSALLOPAEAPAVEENVTETVAETSDPHTA 720
Db 1088 SETGQATPESLEVEFVTDVHVATCQVIKQLQMEQAVAPESSETLTDSETNGSTPLAD 1147
Qy 721 ADDLSALLOPAEAPAVEENVTETVAE-----IPDFNATADDLSALLOPSEVPAVENAEIT 777
Db 1148 SDTAGTQDETIDSDQSKATAAQRQSVTEEEAATAQKEBPSTLPNNVPAQEEHGEPP 1207
Qy 778 LET--PDSNTSEADALPDFLKD--GERETVDWSIYLSEENI-----PNN---AD 819
Db 1208 RDVLEPTQCELTAAPVPLAKTEVQGEVDW-----LGEKVKEQEVFVHSGFNSQKAD 1264
Qy 820 TSFPESEVSG-SDAPSEAKYDLAEMYLEIGDRDAAAE-----TVQKLEAEAB 864
Db 1265 VTVDSEVNGVAGCQEKESTEVQSLSLEGEEMTIDVEKEKRETKPEQVSEGE 1316

RESULT 11
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match          5.4%; Score 236; DB 9; Length 2478;
Best Local Similarity 21.5%; Pred. No. 7.8e-06;
Matches 188; Conservative 137; Mismatches 326; Indels 224; Gaps 41;

Qy 46 ASVAVASQAQAHAGLGLNATQSNLDEFFSGSIITVGBEAKALLGGSVTVSEKGLTAKVH 105
Db 757 ATTAKAAALEEFDEV---VQAIDQAPLNPDTTNEEVAEAI-----ERINAAKVS 803
Qy 106 KLGGKAVIASSQAVRDPVLVFERIGAGQVRVETAILDPVGYSPKTKSALSQKTHRKT 165
Db 804 --GVKAIEAVTTAQDLE-----RV-----KNEISKIENITDSTQTK---MAYNEVKQ 847
Qy 166 APTAESQENQAKALRKTDKK-DSANAAPKPYNGKTHTVRKGETVKQIAAARPHLTL 224
Db 848 AATA--RKAQNAVTSNATNEEVAEADAAVDAQKQGLHDI---QVVK-----SK 891
Qy 225 EQVADALLKANPNVSAHGRLRAGSVLHIPNLRIKAEOPKQOTAKPKAE-----TAS 276
Db 892 QEVADTKSKV-----LDKINAIQTQAK-VKPAADTEVENAYNTRK 930
Qy 277 MPSEPSQATVEKPEKPEAKVAPEKPAVRPEPVAANTAASETAESAPQAAAA 336
Db 931 QEIQNSNASTE-----EKQAYTELDTKKQEARTN-LDAANTNSVITAKD-NSIAAI 982
Qy 337 SAIDTPTDETCNAVSEPVQVSAEEETESGLFGSYTLLAGGGAALIALLLLLRLAQSK 396
Db 983 NQVQAATTKSDAKAE-IAQKASERKTAIEAMNDSTI-----EEQQA 1023
Qy 397 RAARTEBSVPEERPDLDAAADDG-----IETFAEVETPATPEPAPKNDVNDTL--- 445
Db 1024 AKDKVDQAVVTANADIDNAAANNDVDNAKTNTNEATIAAITPDANVKPAAQAIADKVOAQ 1083
Qy 446 --ALDGE---SEELSAKOTFVETDTPSNRIDLDFDSLAAAQNGILSGALTODEETQK 499
Db 1084 ETAIDGNGSTTEKAAAKQOVTEKTTADAAD-----AAHTNA-----EVEAAK 1129
Qy 500 RADADWNAIBSTDSVYEPETFPNPNVBEIVIDTPESVSAQTAENKPEKTVDTDFSDNLP 559
Db 1130 KA-----AIAKIEAI-QPATTTKDNKE-AIATKANERKTAIAQTODITABEIAAANADV 1182
Qy 560 NNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVQTPPELHDFLKVYETDAVE 619
Db 1183 DNAV-TQANSNIEAAN-----SONDVQAKTTGE-----NSIDQ 1215
Qy 620 TARETPDFNAAADLSALLOPAEAPSVEENITETVAETPDNFNATADDLSALLOPSEVP 679
Db 1216 VTPVNVKATARNETAIL-----NNKLOEIQTAPD--ATDEE---KQADAEAN 1260
Qy 680 EENA---AEIVADDLSALLOPAEAPAVEENVETVAETSDPHTAADLSALLOPSEVP 736
Db 1261 TENGKANQAISAAITNAQVDEAKANA-EAALNAVTPKVVKQAQAKDEID-----QLQAT 1313
Qy 737 EENTKTVAEIPDFNATADDLSALLOPSEVPVAENAEITLETTPDSNTSEA-DALPDFL 795
Db 1314 QTNVINN-----DQNTATTEKEAAIQ-QLATAVTDAKNNITPAATDDNGVDQAKQAGKNSI 1367
Qy 796 KDGREETVDWSIYLSSEENIPNNADTSFPPSEVSGSDAPSEAKYDLAEMYLEIGDRDAAAE 855
Db 1368 QSTOPATA-----VKSNKNDVDQAVTTQNCQAINDTTGTATTE-----EKNAAKDL 1412
Qy 856 VQKLEEAEGDVLK-----RAQALAEIGI 880
Db 1413 VLKAKKAYQDILNAQTNDVTQIKDQAVADIQGI 1447
```

RESULT 12

```

US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Carrick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12967
; LENGTH: 2478
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match          5.4%; Score 236; DB 9; Length 2478;
Best Local Similarity 21.5%; Pred. No. 7.8e-06;
Matches 188, Conservative 137; Mismatches 326; Indels 224; Gaps 41;

Qy      46 ASVAVASFOAHAGLGLNLTQSNLDPEFGSITVTGEEAKALGGSGVTSEKGLTAKVH 105
Db      757 ATTAKAAALEBFDEV---VQAQIDQAPLNPDTTNEEVAEAI-----ERINAAKVS 803

Qy      106 KLGDKAVIANSSEQAVRDPLVLFPIGAGAVREYTAILDPVGYSPKTKSALSODKTHRT 165
Db      804 --GVKALEATTTAQDLE-----RV-----KNEEISKIENTDSTQTK---MDRAYNEVKQ 847

Qy      166 APTAESQENQAKALRTDKK--DSANAAVKPAYNGKTHTVKRGKTVKQIAAAIRPKHLTL 224
Db      848 AATA--RKAQNAVTSNATNEEVAEADAADAAQOQLHDI---QVVK-----SK 891

Qy      225 EQVADALLKANPNVSAHRLRAGSVLHPLNLRIKAEQPKPQTAKPAE-----TAS 276
Db      892 QEVADTKSKV-----LDKINAIQTQAK--VKPAADTEVENAYNTRK 930

Qy      277 MPSEPSQATVEKEPEVEKPAKVAPEAKPAVRPEVPVAANTAASETAESAPQAAAA 336
Db      931 QEIQNSNASTE-----EKQAAYTELDTKKQEARTN--LDAANTNSDVTAKD--NSIAAI 982

Qy      337 SAIDPTPTDETCNAVSEPEVQSVASEETESGLFGGSGYTLILLAGGGAALIALLLRLAQSK 396
Db      983 NQVQAATTKKSDAKAE--IAQASERKTAIEMNDSTT-----EEQQA 1023

Qy      397 RARTESVPEEPDLDAAADDG-----IEITFAEVETPATPEPAPKNDVNDTL--- 445
Db      1024 AKDKVDQAVVTANADIDNAAANNNDVNAKTTNEATIAAITPDANVPAKQAIADKVQAO 1083

Qy      446 --ALDGE---SREELSAKQTFDVETTPSNRIDLDFDSIAAQQNGILSGALTDEETQK 499

```

```
Db 311 QSRGGQQRQSRIDAVPASTQEQDQA-----QSSSPAPAGSSEBEGVGSTWE 358
Qy 437 -----PKNDVNDTLALDGESEBELSAKOTFVETDTP--SNRID--LDFDSLAAQONG 485
Db 359 SFKRLVTPRKKSJKL-----EKEAGRTLTVGACPLRSNRKVENLGFPLFNSSPDG 411
Qy 486 ILSGALTQDEETQKRAADWNAIESTDSVVEPEFTFNPYNPVEIIVDTPEPESVAQTAEN- 544
Db 412 GRGQMGQRQDQA-----TVDSGPEINEDPDPAVVPVPLSEYD 450
Qy 545 --KPETVDTDFSDMLPS-----NNHI-----GTEETASAKPASPSGLAGF 582
Db 451 AVEREKMAQGNALPSCWGVSEELSKTLVHVSVAVIDGTTRAVTSVEERSPSWISAS 510
Qy 593 L-----KASSP-ETILEKTVAEVQTP-----BELHDFLKVYETDAVAE--TA 621
Db 511 VTEPLEHTAGEAMPVEVEVTEKDIABETPVLTOTLPEGKDAHDMVTSEVDFTSEAVTA 570
Qy 622 PETPD-----FNAADDSLALQPAEAPSVENITET-----VAETPD----- 659
Db 571 TETSEALTRTEVTEASGABETDMVSAVSQLTSDPDTEETATPVQEVGGVLDTEEREQ 630
Qy 660 ----FNATADDLSALQPSVPA-----VEENA----- 683
Db 631 TQAILQAVADKVK---RESQVPATQTVQRTGSKALEKVEEVEDSEVLASEKEKDVMPKG 687
Qy 684 ----AEIVAD-----DLSSALQPAEAPVENV 707
Db 688 PVQBAGAEHLAQSGSETGQATPESLEVPEVTADVHVATCOVIKLQOLMEQVAPESSETL 747
Qy 708 TETVAETSDFHTAADDLSALQPAEVPVAVENVTKTVAE---IPDENATADDLSALQPS 764
Db 748 TDSSTNGSTPLADSDTADGTQDFTIDSQSKATAAARQSOVTEERATAQKEPSTLPN 807
Qy 765 EVPAVEENAABITLET--PDSNTSEADALPDLKD--GEETVDWMSIYLSSENI----- 814
Db 808 NVPAQEEHGEPRGSDVLEPTQELAAAVPVWQKTEVGQGEVDW---LDGEKVKEQEV 864
Qy 815 -----PNM---ADTSFSESVG--SDAPSEAKYDLAEMVLEIGDRDAAE-----TVQKL 859
Db 865 FVHSGSPNSKAADVTYDSEVMGVAGCQKESTEQSLSEGERMETDVEKEKRETKPEQV 924
Qy 860 LEEAE 864
Db 925 SERGE 929
```

RESULT 14

US-09-839-996-6

Sequence 6, Application US/09839996

Publication No. US20030009010A1

GENERAL INFORMATION:

APPLICANT: St. Gene III, Joseph W.

Falkow, Stanley

TITLE OF INVENTION: Haemophilus Adherence and Penetration

Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/839,996

FILING DATE: 20-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791

FILING DATE: 25-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6

Query Match 5.3%; Score 232.5; DB 11; Length 1848;

Best Local Similarity 19.2%; Pred. No. 8.6e-06;

Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

```
Qy 35 LKNRRIKLIASAASVAAAFQAHAG-----LGLNI-QSNLDPPFGSGTIVT 80
Db 766 LYSGRNVANITSNITASNNAQVHIGYKTDGTVCVRSYDTGYVTCHNSNLSEKALNSFNPT 825
Qy 81 GEEAKA-LLGGSGVTYSEKGLTAKVHKLGDKAVIATVSEQAQVAVRDPVLVFRIGAGAQAQVREY 139
Db 826 NLRGNVNLTENASFTLGLKANLFGTIQSIGTSQVNLKENS-----WHLTGSNVNQL 877
Qy 140 TAILDPVGYSPKTSALSQDKTRKTAPTAESEQENQNAKALRK-----TDCK 186
Db 878 N-----LTNGHILHNAQNDANKVITYNTLTVNSLSGSGSYFYWDFNNK 922
Qy 187 DSANAAVKPAYNGKHTV--RKGETVKQIAAAIRPKH--LTLEQVADA-----LLKANP 236
Db 923 SNKVVVNSKATGNTFLQVADKTGE-----PNEHNLTLFASNATRNLEVTLANG 972
Qy 237 NV-----SAHGR-----RAGSVLHPIPLNRIKAEQPKQTAKE---PKA 272
Db 973 SVDRGAWKYKLRNVNGRYDLYNPEVEKRNQVTDTNITTTNDIQADAPSAQSNNEEARV 1032
Qy 273 ETASMPSEPSKQATV--EKPVKEDEAKVAAPAEAKAPVAPRPPVPAANTAA----- 322
Db 1033 ETPVPPAPATESAIAEQPETRP-AETAQAPAMESTNTANSTETAPKSDTATOTENPSE 1091
Qy 323 ---SET-----AASAPQEAASAIIDTPTDETGNVAVSPVEQVS 358
Db 1092 SVPSETTEKVAENPPQENETVAKNEQATEPTQNGEVAKEDQPTVEANTQTNEATQSEG 1151
Qy 359 ABEETESGLFGSGYTLILLAGGGAALIALLLLRQAQSKRAARTEESVPEEPD--LDDAA 416
Db 1152 KTEETQT-----AETKSEPTESVTVSENOPEKTVSOST 1184
Qy 417 DDGLEI---TPAEVET-----PATPEAPKNDVNDTLALDGESEBELSAKQ 459
Db 1185 EDKVVVEKEEKAKVETETQKAPQVTSKEPKQAPAPPEVPTDTNABEAQALQQTQPT 1244
Qy 460 TFDVETTPSNRIDLDFSLAAQNG-----ILSGALTQDEETQKRAADWNAIESTDS 513
Db 1245 VAAAEITSPNSKPAEETQQPSEKTNAPFVTVSVSENTATQPTETETAKVE---KEKTQE 1301
Qy 514 VYEPETENPNPVEIIVDTPEPESVAQTAENKPTVDTDFSDNLPSNNHICTEETASAKP 573
Db 1302 V--PQVASQESPKQ---EQPAKPAQQT---KPQABFA--RENVLTTKNVG-EQPPQAP 1350
Qy 574 ASPS---GLAGFLKASSPETILEKTVAEVQTFPEELHDFLKVYETDAVAETAPETPDFNA 630
Db 1351 QTQSTAVPTTGTETAANSKPAKPAQAKPQEPARENVSTVNTKEPQSQTSATVSTEQPA 1410
Qy 631 ADDLSALQPAEAPSVENITETVAETPDFNATADDLSALQPSVPAVEENAAEIVADD 690
```

```

Db      1411 KETSNVEQAPAPENSINTGSAATTMTET-----AEKSKPQME-----IVTEND 1453
QY      691 LSALLQPAEAPAEVENVETVAETSDPHTAADDLSALLQPAEVPVAVENVTKTVAEIPDF 750
Db      1454 ----RQP-EANTVADNSVANNSESSE--SKGRRRSVSPKETSABETTVASTQ----- 1500
QY      751 NATADDLSALLQPSB---VPAVEENAAEITLETDPDNTSEADALPDFLKDGEETVDWSI 807
Db      1501 ETTVDNSVSTPKPRSRRTSRVSQINSYE-PVELPTENAENAVNQ----- 1544
QY      808 YLSEENIPNNAD-----TSFSESVGSDAPSEAKYDLAEMYLEIGDRDAAAETVOKLLEE 862
Db      1545 --SGNNVANSQPALRNLTSKNTNAVISNAKAPF---VALNVG--KAVSQHISQLEMN 1596
QY      863 AEG 865
Db      1597 NEG 1599

RESULT 15
US-10-080-505-6
; Sequence 6, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1848
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-6

```

```

Query Match      5.3%; Score 232.5; DB 15; Length 1848;
Best Local Similarity 19.2%; Pred. No. 8.6e-06;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

QY      35 LKNRQIKLIAASVAVASFOAHAG-----LGGLNI-QSNLDEPFGSITVT 80
Db      766 LYSGRNVANITSNITASNNAQVHGYKGTGTVCSRSDYTGVTCHNSNLSEKALNSFNP 825
QY      81 GEEAKA-LLGGGSVTVSEKGLTAKVHKGDKAVIAVSSEQAVRDPVLVFRIGAGAQVREY 139
Db      826 NLRGNVNLTENASFTLGKANLFGTIQSIGTSQVNLKENS-----WELTGNNSVNQL 877
QY      140 TAILDPVGYSPKYSALSQDKTHRKTAAPTAESENQNAKALRK-----TDKK 186
Db      878 N-----LTNGHILHNAQNDANKVTTYNTLVNSLSGNGSFYYWVDFNNK 922
QY      187 DSANAARKPAYNGKTHTV--RKGETVKQIAAARPKH--LTLEQVADA-----LLKANP 236
Db      923 SNKVVNKSATGNFTLQVADKTGE-----PNHNELTDFDASNATRNLEVTIANG 972
QY      237 NV-----SAHGRLL-----RAGSVLHPIPNLRKARQPKPOTAK---PKA 272
Db      973 SVDRGAWKYLRVNGRYDLYNPEVEKRNQTVTTNTTTPNDIQADAPSQAQSNNEBIARV 1032
QY      273 ETASMPSEPSQATV--EKPVEKPEAKVAEPAKPAVRPEPVPAAANTAA----- 322
Db      1033 ETPVPPPPAPATESAIASEQETRP-AETAQPAEMEINTANSTETAPKSDTATQTENPSE 1091
QY      323 ---SET-----AESAPQEAASAIDPTTDTGNVASEPVEQVS 358
Db      1092 SVPESETTEKVAENPPOENETVAKNEQATEPTPQNGEVAKEQDQPTVEANTQTNATQSE 1151

```

```

QY      359 AEBETESLFGSYTLLLAGGGAIALIALLLRLIAQSKRARTTESVPEEBPD--LDDAA 416
Db      1152 KTEETQT-----AETKSEPTESVTVENQPEKTVSQST 1184
QY      417 DDGIEI--TFABVET-----PATPEPAPKNDVNDTFLALDGESEELSQAQ 459
Db      1185 EDKVVVEKEEKAKVETRETQKAPQVTSKEPPKQABEPEEVPDTNABEAQALQOTQPTT 1244
QY      460 TFDVETTPSNRIDLDFDSLAAAQNG-----ILSGALTQDEETQKRAADADNAIESTDS 513
Db      1245 VAAAEITSPNSKPAEETQQPSEKINAEPVTPVVSSENTATQPTETEETAKVE---KEKTOE 1301
QY      514 VYEBETPNPNPVEIVIDTPEPESVAQTAENKPKETVDTDFSDNLPSNNHIGTEETASAKP 573
Db      1302 V--PQVASQESPKQ---EQPAAPKPAQOT---KPOAEPK--RENVLTTKNVG-EPQPQAQP 1350
QY      574 ASPS---GLAGFLKASSPETTILEKTVAEVOTPEELHDFLKVYETDAVAETAPETPDFNAA 630
Db      1351 QTQSTAVPTTGETAANSKPAKPAQAKPQTEPARENVSTVNTKEPOSQTSATVSTQPA 1410
QY      631 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPSVPAVEENAAEIVADD 690
Db      1411 KETSSNVEQAPAPENSINTGSAATTMTET-----AEKSKPQME-----TVTEND 1453
QY      691 LSALLQPAEAPAEVENVETVAETSDPHTAADDLSALLQPAEVPVAVENVTKTVAEIPDF 750
Db      1454 ----RQP-EANTVADNSVANNSESSE--SKGRRRSVSPKETSABETTVASTQ----- 1500
QY      751 NATADDLSALLQPSB---VPAVEENAAEITLETDPDNTSEADALPDFLKDGEETVDWSI 807
Db      1501 ETTVDNSVSTPKPRSRRTSRVSQINSYE-PVELPTENAENAVNQ----- 1544
QY      808 YLSEENIPNNAD-----TSFSESVGSDAPSEAKYDLAEMYLEIGDRDAAAETVOKLLEE 862
Db      1545 --SGNNVANSQPALRNLTSKNTNAVISNAKAPF---VALNVG--KAVSQHISQLEMN 1596
QY      863 AEG 865
Db      1597 NEG 1599

```

Search completed: December 12, 2003, 17:38:16
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:29:17 ; Search time 26 Seconds
(without alignments)

3254.940 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCPPMTKFTDCT.....EEABGDVLKRAQALAEILGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76.*

1: Pirl1.*

2: Pirl2.*

3: Pirl3.*

4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4404	100.0	880	2 G81786	Neisseria-specific
2	4273.5	97.0	875	2 C81209	TspA protein NMB03
3	381.5	8.7	919	2 F83257	hypothetical prote
4	353.5	8.0	1621	2 A82255	hypothetical prote
5	316	7.2	1822	2 S33441	EF protein - Strep
6	312	7.1	801	2 T29018	hypothetical prote
7	276.5	6.3	1262	2 T22523	hypothetical prote
8	272.5	6.2	1829	2 T24583	hypothetical prote
9	271.5	6.2	1110	2 I51116	NF-180 - sea lampr
10	269	6.1	2364	2 A56577	microtubule-associ
11	266.5	6.1	1634	2 T26517	hypothetical prote
12	264	6.0	2055	2 T31110	extracellular matr
13	263.5	6.0	2464	1 QRMSP1	microtubule-associ
14	260.5	5.9	880	2 D89756	protein T2387.2b
15	258.5	5.9	5327	2 T13564	microtubule-associ
16	257	5.8	1337	2 T30291	dextranase - Strep
17	253.5	5.8	682	2 F83228	hypothetical prote
18	253.5	5.8	3507	2 T34513	hypothetical prote
19	253	5.7	1367	1 S48478	glucan 1,4-alpha-g
20	248	5.6	873	2 A47283	hypothetical prote
21	248	5.6	971	2 T19431	calphotin - fruit
22	244.5	5.6	734	2 B42680	hypothetical prote
23	244.5	5.6	2570	2 T17451	nucleolus-cytoplas
24	244	5.5	1229	2 T25697	fimbriae-associate
25	244	5.5	1616	2 G64242	hypothetical prote
26	243.5	5.5	1558	2 B71603	cytadherence-acces
27	243	5.5	865	2 A47282	RESA-H3 antigen PR
28	241	5.5	2187	2 T30826	calcium-binding pr
29	240.5	5.5	1038	2 AG2187	nascent polypeptid
					hypothetical prote

30	239	5.4	1664	2 T18262	S-layer protein -
31	237.5	5.4	506	2 S47439	I2 protein - Trypa
32	236.5	5.4	1983	2 AC1922	two-component hybr
33	236	5.4	6642	2 T29757	protein UNC-89 - C
34	235.5	5.3	2484	2 T26216	hypothetical prote
35	235.5	5.3	2607	2 T26215	hypothetical prote
36	235	5.3	1320	2 JC5630	TCOF1 protein - mo
37	235	5.3	1346	2 A57376	probable regulator
38	233.5	5.3	1230	2 T22458	hypothetical prote
39	232.5	5.3	1849	2 C41859	IGA-specific metal
40	232	5.3	1200	2 A46194	neurofilament prot
41	230	5.2	1072	1 A37221	neurofilament trip
42	229.5	5.2	2481	2 D90011	FntB protein [impo
43	227.5	5.2	2472	2 E83594	still frameshift p
44	225.5	5.1	1046	2 T29776	hypothetical prote
45	223.5	5.1	3942	2 T42730	Bassoon protein -

ALIGNMENTS

RESULT 1

G81786

Neisseria-specific antigen protein, TspA NMA2146 [imported] - Neisseria meningitidis (str G81786)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell ; Holroyd, S.; Jagele, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

C:Accession: G81786

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:2022556; PMID:10761919
A:Accession: G81786

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-880 <PAR>

A:Cross-references: GB:A1162758; GB:A1157959; NID:g7380672; PIDN:CAB85358.1; PID:g738076;

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: tspA; NMA2146

Query Match 100.0%; Score 4404; DB 2; Length 880;
Best Local Similarity 100.0%; Pred. No. 1.6e-194;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGRLPRRCPPMTKFTDCTRSNRIOPTTHRGYILKNNRQIKLIAASVAVASQAHAAGL 60

Db 1 MPAGRLPRRCPPMTKFTDCTRSNRIOPTTHRGYILKNNRQIKLIAASVAVASQAHAAGL 60

Qy 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120

Db 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120

Qy 121 VRPVLVFRIGAGAVREYTAILDPVGYSPKTSALSDGKTHRTAPTASQENQNAKAL 180

Db 121 VRPVLVFRIGAGAVREYTAILDPVGYSPKTSALSDGKTHRTAPTASQENQNAKAL 180

Qy 181 RKTDDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAAIKPKHLTLEQVADALLKANPNVSA 240

Db 181 RKTDDKDSANAAVKPAYNGKTHTVRKGETVKQIAAAAIKPKHLTLEQVADALLKANPNVSA 240

Qy 241 HGRIRAGSVLHIENLRKIAEQPKPTAKPKAETASMPSEPSKQATVEKPEKPEAKVAA 300

Db 241 HGRIRAGSVLHIENLRKIAEQPKPTAKPKAETASMPSEPSKQATVEKPEKPEAKVAA 300

Qy 301 PEAKPEKPAVRPVPVPAANTAASETAESAPOAAASAIPTDPTDTCNAVSEPVQVSAE 360

Db 301 PEAKPEKPAVRPVPVPAANTAASETAESAPOAAASAIPTDPTDTCNAVSEPVQVSAE 360

Qy 361 EETESGLFGGYSYTLILLAGGGAALIALILLRLAQSKRARTEESVPEEPDLDAAADGI 420

Db 361 EETESGLFGGYSYTLILLAGGGAALIALILLRLAQSKRARTEESVPEEPDLDAAADGI 420

QY 421 EITFAEVETPATPEPAPKNDVNTLALDGESEBELSAKOTFDVETDTPSNRIIDLDFSLA 480
DB 421 EITFAEVETPATPEPAPKNDVNTLALDGESEBELSAKOTFDVETDTPSNRIIDLDFSLA 480
QY 481 AAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNVVEIVDTPEPSVAQ 540
DB 481 AAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNVVEIVDTPEPSVAQ 540
QY 541 TAENKPBETVDTSDNLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
DB 541 TAENKPBETVDTSDNLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
QY 601 TPPELHDFLVYETDAVAETAPETPDNAAADDLSALLOPAEAPSVENETETVAETPDF 660
DB 601 TPPELHDFLVYETDAVAETAPETPDNAAADDLSALLOPAEAPSVENETETVAETPDF 660
QY 661 NATADDLSALLOPSEVPAVEENAAEIVADDSALLOPAEAPAVENVTETVAETSDPHTA 720
DB 661 NATADDLSALLOPSEVPAVEENAAEIVADDSALLOPAEAPAVENVTETVAETSDPHTA 720
QY 721 ADDLSALLOPAEAPAVENVTETVAETPDNATADDLSALLOPSEVPAVEENAAEITLET 780
DB 721 ADDLSALLOPAEAPAVENVTETVAETPDNATADDLSALLOPSEVPAVEENAAEITLET 780
QY 781 PDSNTSEADALPDFLKDGBEETVDSIYLSEENIPNNADTSFPSESVDGSDAPSEAKYDLA 840
DB 781 PDSNTSEADALPDFLKDGBEETVDSIYLSEENIPNNADTSFPSESVDGSDAPSEAKYDLA 840
QY 841 EMYLEIGDRDAAAEFTVQKLEAEAGDVLKRAQAQAEELGI 880
DB 841 EMYLEIGDRDAAAEFTVQKLEAEAGDVLKRAQAQAEELGI 880

RESULT 2
C81209
tspA protein NMB0341 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81209
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigiani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81209
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-875 <TET>
A:Cross-references: GB:AB002391; GB:AB002098; NID:g7225561; PIDN:AAF40784.1; PID:g722556
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0341

Query Match 97.0%; Score 4273.5; DB 2; Length 875;
Best Local Similarity 97.1%; Pred. No. 1.5e-188;
Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

QY 1 MPAGLRPRCPMWTFTDCTSRNIQPPPTHERGYILKNNRQIKLIAASVAAASFQAHGL 60
DB 1 MPAGLRPRCPMWTFTDCTSRNIQPPPTHERGYILKNNRQIKLIAASVAAASFQAHGL 60
QY 61 GGLNIQSLDLPFGSGITVTEEAQKALGGSVTVSEKGLTAKVHKLGDKAVIASSQQA 120
DB 61 GGLNIQSLDLPFGSGITVTEEAQKALGGSVTVSEKGLTAKVHKLGDKAVIASSQQA 120
QY 121 VRDPVLFRICAGAVREYTAALDPVGSPTKSGALSQKTRKTAPTAEQENQAKAL 180
DB 121 VRDPVLFRICAGAVREYTAALDPVGSPTKSGALSQKTRKTAPTAEQENQAKAL 180
QY 181 RKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAIPKHLTLEQVADALLKANPNVSA 240

DB 181 RKTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAIPKHLTLEQVADALLKANPNVSA 240
QY 241 HGLRAGSVLHIPLNRIKAEQPKQTAKEPASTMPSEPSKQATVKEKPEAKVAA 300
DB 241 HGLRAGSVLHIPLNRIKAEQPKQTAKEPASTMPSEPSKQATVKEKPEAKVAA 300
QY 301 PEAKAEKPAVRPEPVAANTAAETAESAPOEAAASAIPTDPTDTCGNVSEPVQVSAE 360
DB 301 PEAKAEKPAVRPEPVAANTAAETAESAPOEAAASAIPTDPTDTCGNVSEPVQVSAE 360
QY 361 EETES-----GLFGSGYTLILAGGGAIALLLRLAQSRRARTESVPEEPDLDAA 416
DB 361 EETESGLFDGLFGSGYTLILAGGGAIALLLRLAQSRRARTESVPEEPDLDAA 420
QY 417 DDGIEITFAEVETPATPEPAPKNDVNTLALDGESEBELSAKOTFDVETDTPSNRIIDLDF 476
DB 421 DDGIEITFAEVETPATPEPAPKNDVNTLALDGESEBELSAKOTFDVETDTPSNRIIDLDF 480
QY 477 DSLAAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNVVEIVDTPEPE 536
DB 481 DSLAAQNGILSGALTQDEETQKADADWNAIESTDSVYEPETPNPNVVEIVDTPEPE 540
QY 537 SVAQTAENKPBETVDTSDNLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTV 596
DB 541 SVAQTAENKPBETVDTSDNLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTV 600
QY 597 ABVQTPBELHDFLVYETDAVAETAPETPDNAAADDLSALLOPAEAPSVENETITVAE 656
DB 601 ABVQTPBELHDFLVYETDAVAETAPETPDNAAADDLSALLOPAEAPSVENETITVAE 660
QY 657 TPDFNATADDLSALLOPSEVPAVEENAAEIVADDSALLOPAEAPAVENVTETVAETSD 716
DB 661 TPDFNATADDLSALLOPSEVPAVEENAAEIVADDSALLOPAEAPAVENVTETVAETSD 720
QY 717 FHTAADDLSALLOPAEAPAVENVTETVAETPDNATADDLSALLOPSEVPAVEENAAE 776
DB 721 FNATADDLSALLOPSEAPAVENAAETV-----ADDSALLOPAEAPAVENAAE 771
QY 777 TLETPDNTSEADALPDFLKDGBEETVDSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 836
DB 772 TLETPDNTSEADALPDFLKDGBEETVDSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 831
QY 837 YDLAEMYLEIGDRDAAAEFTVQKLEAEAGDVLKRAQAQAEELGI 880
DB 832 YDLAEMYLEIGDRDAAAEFTVQKLEAEAGDVLKRAQAQAEELGI 875

RESULT 3
F83257
hypothetical protein PA3115 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83257
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-919 <STO>
A:Cross-references: GB:AB004735; GB:AB004091; NID:g9949216; PIDN:AAG6503.1; GSPDB:GN00
C:Genetics:
A:Experimental source: strain PA01
A:Gene: PA3115

Query Match 8.7%; Score 381.5; DB 2; Length 919;
Best Local Similarity 23.0%; Pred. No. 3.6e-10;
Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;

QY 41 IKLIAASVAAASFQAHG-GLGGLNIQSLDLPFGSGITVTEEAQKALGGSVTVS--- 96

```

Db      8 VRAIAAA-SVLTSCWAHGLGELITLSALNQFLDAEILL--EVRD-LGSGEVIPLSAS 63
Qy      97 -EKGLTAKVHKL-----GDKAVIAYSSEQAVRDPVLVFRIGA---GAQVR 137
Db      64 PEEFSKAGVDRLYLTLTKFTPVVKPKNGSKVIRTSKPKQEPYLFVQVLPWNGRLLR 123
Qy      138 EYTAILDPVGYSPKTKSALSOGKTHRTAPTAESENQNAKALRKTDKUSANNAVK--- 194
Db      124 EYTVLLDPPIYSPOAAASAPQPV---SAPRATG-----APRAPQAPAVRTTA 169
Qy      195 PAYNGKTHTVRKGETVTKQIAAAIRPKHLTLEQVADALLKANPNVSAHG---RLRAGSVLH 251
Db      170 PAGSDTYRTV-SNDTLWEIAQBNETDRVSPQAMLAFOELNPGAFVGNINRLKASGVLR 228
Qy      252 IPNLNRIKAPQPKQTAAPKPAETAS-----MPSEPSK-QAT 286
Db      229 IPTEQMLERSPREALSQVQONQSWGRSPNPAAGSAGARQLDATORNAAGSAPSKVDAT 288
Qy      287 ----- 286
Db      289 DNLRLVSGEKASKGADKGGKGSKAITADTLAVTKESLDSTRNEELQRMQDLQSLD 348
Qy      287 -VEKPEKPEAKVA-----APEAKAEKPAVRPEPVPAANTA 321
Db      349 KLQKLIQLKDAQLAKLOGLGAGGGAAPNAALPDASQPNAAQAQAPQCTPAAAPTP 408
Qy      322 ASETAASAPQEAASAIIDPTDE--TGNVSEPVQVSABEETESGLFGGSYTLILAG- 378
Db      409 APAGEAPAAPQPVVAPPAPAAEKPPAPAPAPAPVQAAEQAPAPSEF-----DELLANP 463
Qy      379 -----GGAALIALLLLRLAQSKRARTESVP-----EEBPLDDAADGI-BITF 424
Db      464 LWLAVIGSALLLVLLMLSLRNNAQKEEAQFAADTGEQEDALDLKGQFDDUTL 523
Qy      425 ARVETPATPPAPKNDVNTLALDGESEEL-----SAKOTFOVETDTPSNRDLDD 475
Db      524 DEPE-POVAVAPOVEKTTAQTSDALGADIIYIAGRFNQAAELLQNAIYDEP-QRTDLR 581
Qy      476 FDSLAA-AQNGILSGALTQDEETQKADADNAIESTDSVYEPFTFPNPNPVEIVIDTPE 534
Db      582 LKLMVVAEMGDREGFARQENELREIGGAQ-POVEQLKRY-----PAMVAVAAVA 631
Qy      535 PESVAQTAENKPEVTDVDFSDNLPNNHIGTEETASAKP-ASPSGLAGFLKASSPETILE 593
Db      632 GLAGAKLAQDELQSFSLD-DLSLDDSGH-----AAKPAAGQDLDDAFQSLDLDLGGD 683
Qy      594 KTVAEVQTPR-ELHDFLKVYETDAVETAPETPDENAAADDLSALLQAPABAPSVEENITE 652
Db      684 DVQADLAKSDGALDDLLDSDLLAASTPADKP-----VDLLDFGLDFAE----- 728
Qy      653 TVAETPDENATADLSALLQSEVPVAVENNAEIVADDLSALIQ---PAPAPAVEENVTE 709
Db      729 -LAETPS-QPKHDDLQGFSLDLOAP--EDKLSD---DDFLSLINDEVPAAAPADNEFTLD 781
Qy      710 TVAETPSDFHAAAD--DLSALLQPAEVPVAVENVTKTVAEIPDFNATADDLSALLQSEVP 767
Db      782 TEAAEPEALSLPDDFSLADEPTEPAPEKGEFSFAAQQLDEVSQAQDELAS----- 833
Qy      768 AVEENAAEITLETPSDNT---SEADALPDLFKGEEETVDWSIYLSSEENIPNNADTSFPS 824
Db      834 -----NLDEPKSATPSFAEDAASALDGD-----ADDDDFD 866
Qy      825 ESVGSDAPSAKYDLAEWYLEIGORDAAAETVQKLLBEAGDVILKRAQ 873
Db      867 LSGADEAAT--KLDLARAYIDMGDSEGARDILDEVL--AEGNDSQQA 911

```

RESULT 4

A82255
 hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C;Accession: A82255
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A;Reference number: A82035; MUID:20406833; PMID:10952301
 A;Accession: A82255
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1621 <HEI>
 A;Cross-references: GB:AE004181; GB:AE003852; NID:G9655454; PIDN:AAF94159.1; GSPDB:GN0012
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VC0998
 A;Map position: 1

Query Match 8.0%; Score 353.5; DB 2; Length 1621;
 Best Local Similarity 23.1%; Pred. No. 1.5e-08;
 Matches 226; Conservative 129; Mismatches 316; Indels 307; Gaps 50;

```

Qy      137 REYTAILDPVGYSPKTKSALSOGKTHRTAPTAESENQNAKALRKTDKUSANNAVKPA 196
Db      24 RFFORLLLPVAVMVVVTQTSFVSAESIRLVGPDGQVQPTPQ---YSENIVRNSANN--EPG 78
Qy      197 -YNGKTHTVRKGETVTKQIAAAIRP-KHLTLEQVADALLKANP-----NVSAHGRLRAGSVL 250
Db      79 RFFGPTSA---NOTLMSIASQLRPSSTVVOOTLLAIYQLNPQAFENQNTHT-TLIPSTL 134
Qy      251 HINLNRKIAEQPKPQTAAPKPAETASMPSEPSQATVEKPEKPE-AKVAAP--EAKAEK 307
Db      135 RVPSLAQISNSQDQVAVINASHQAKLNQTPD---TVRPVAPPRPAPVATPKVEVATQ 191
Qy      308 PAVRPEVPVPAANTAASETAASAPQEAASAIIDPTT----- 343
Db      192 P---PQVTPT-----TAQEKAPTTELKTPAKPSQSDAEVMALEEKNHTLRML 237
Qy      344 -----DETG--NAVSEPVQVSABE---ETESGLFGGSYTLILAGG----- 380
Db      238 SQVQSEVSTLKEGLDENRIRSEVERLLEERKAEASRLAFSALDNLNSGMLVALLA 297
Qy      381 -----AALIALLLLRLAQSKRARTESVPREEDLDDAADGIBITAEVETPATPE 434
Db      298 LIFGLLIAVVLNLRSSAQOENPTQNNITEMP-----TAAFTVLG 341
Qy      435 PAPKNDVNDTLALDG-----ESEEELSAKQTF---DV-----ETDTPSN----- 470
Db      342 PEQTEDIGDILLDDDLFSTTDDKEENDAEKAFSDEDDVFADLNETDLDNFNLQDQSDDL 401
Qy      471 -----RIDLDFDSLAAQNGILSGALTQDE-----ETQKRA-----DADWNAI 508
Db      402 FVGIDDDGDLDTDFDALNESANGI---SVNADDKALGLEEMERALNDVSBPTDNDLNSF 458
Qy      509 ESTDSVYEPFTFPNPNPVEIVIDTPEPE-----SVAQTAENKP----- 546
Db      459 DLADENQMSF-----DDIEALLSGDENELLSPGKVDQSLDLDLASELDALDEPAIQD 513
Qy      547 -ETVTDVDFSDNLPNNHIGTEETASAKPSPSGLAG-----ELHDFLKVYETDAVETA---PETPDFNAAA 631
Db      514 TETLDTLLNDELASL---SEEDDDREFDLGAGVAGQDLDLFASTEEQADLEQLEAKA 569
Qy      586 SSPETILEKTVAEVQTP-----EELHDFLKVYETDAVETA---PETPDFNAAA 631
Db      570 IDETALLDETLAEQDAPLSEESTELDELDDFKPENEDEFDAQADLLOQPEEPILDEE 629
Qy      632 DDLSSALLQ-----PAEAPS---VEENITETV-----AETPDFNATADDLSA 669
Db      630 DSTQLLNEVLGEPVPEELASGLGIDQNSTELLDELDDLDLDDSETEATEFSVAPEKLSV 689
Qy      670 -----LLQPSVPAVEENAAEIVADD-----LSALLQPAEA-----PAVEENV 707
Db      690 EDGTETLFDLELLETEQHPESAESLPPELATEDEFNSDFIDLLNSAPAKOPLLBFFVLDENE 749

```


NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: F51116
R:Jacobs, A. J.; Kamholz, J.; Salzer, M. E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation
X:Reference number: F51116; MUID:95287814; PMID:7770000

[illegible]

66 QSNLDPPFGSITVTGERAK-----ALLGGGSVTVSEKGLT-----AKVHKLGDKAVTA 114
QY : : : : :
177 : : : : :
178 : : : : :
179 : : : : :
180 : : : : :
181 : : : : :
182 : : : : :
183 : : : : :
184 : : : : :
185 : : : : :
186 : : : : :
187 : : : : :
188 : : : : :
189 : : : : :
190 : : : : :
191 : : : : :
192 : : : : :
193 : : : : :
194 : : : : :
195 : : : : :
196 : : : : :
197 : : : : :
198 : : : : :
199 : : : : :
200 : : : : :

Qy 115 VSSEQAVRDPVILFRIGAGAQVREYTAILDVPVGYSPKTKSALSD--GKTHRKTAPTAES 171

QY 172 QENQ-NAKALRKTDKKDSANAAVKPAYNGKHTVRK-----GETVKQIAAA 216
|:::|

DB AEEWFGKFSQIIIEAAEQNDALFSAKEEIEHKRNQKRCLEEDALAGINIEGELNQQSOL 319

QY 290 217 IRPKHILT-----LEQVADAL-----LKANPNVSAHGRLRAG 247

DB 350 MEERHQSDVGNLQDAAQQLENELNENTRWEMARHLREYQDNLNVKMLDIBIAAYKLLJD 409

QY 248 SVLHHPNLNRIKAEQPKPQAKFAETASMPSEPSKQATVEKPV--EKPEAKVAAPKA 305

Db 410 BEIRY-----SSGFLPTAKP-----PKAPSAKPAKVEKKVSKPEIKVES----- 453

Ov 306 EKPVRPEPVA-ANTAAGETAESAPOEAAASAIIDTPTDETGNVSEPVQVSAEEETE 364

[illegible]

Db

```
507 E-----EDRGRKEGEAAEEERAAEEVEVKEEABE--- 535
```

db 536 AEVE-----EAEAEETEAEEAEEEEAAEGEEAAEEVEEEEAKAE-- 585

QY : : : : :
485 GILSGAUHQBEIQKADWNNALISVIEEFIRNFVVDLITLDGVAHML 514
DB : : : : :
586 --AAEAKAEVEEEAABEB--EEGAEEVEEAET---KEEVRAAEVEEGFAEE 636

QY	545 KPEVTDTDFSDNLPNNHIGTEETASAKFASPGLKASSPTELKK---
	601
Dd	637 EAEFEAAE-----EEVTSK-----KAKTQEAEVEEEEEAAAAEA
	672

602. PEELHDFELKVYETDAVETAPETPDFNAAADDLSALLQPAEAPSV--EENITETVETATPDF 660 QY

673 EBEAEE--EAGDEDVEAAEKEEEDSEDSKEAD---AEDEAAEEVEEKEEVTKSDAEAEA 727 DB

QY 661 NATADL SALLQPSVPAVEENAAEIVADDLSALLQPAEAPAVEENVUTETVAETSDHTA 720

QY 721 ADDLSALLQPAEVPAAVEENTKTVAEI-----PDFNATADDLGALLQPEVP 767

QY 768 AVENAAEITLTETPDNTSADALPDFLKDGEETVDWSIYLSEENIPNNATSPFSESV 827

408 EEPDLDDAADDGGIETFAEVETP-----ATPEPAPKNDVNDTLALDGESE 453
1718 ---SAKDGADDLKLSESRPESTTQSKAGSIKDEKSEASRPASVAESVKDEAEKSK 1774
454 ELISAKOTFDVETDTPSNRIDLOFDSIAAQAQNGTLSCALTDDEETQKRADAMNAIESDTS 513
1775 EESRRRESVAESKPLPSK---EASRPASVAESIKDEAKSKESRRRESVAESKPLPS--- 1827
514 VYEPETFPNPNVEIVIDTPEP-----ESVAQTA-----ENKPTVDTDFSDNLP SN 560
1828 ---KEASRPASVAESIKDEAKSKESRRRESVAESKPLPSKEASRPASVAESIKDEAKS 1894
561 NHIGTETASAKPASPSGLAGFLKASPETILKTVAEVOTPELHDFLKVETDAVET 620
1885 KEESRRRESVAESKPLPSK---KEASRPASVAESIKDEAKSKE-----ESRRRESVAEK 1933
621 AP-ETPDFNAAADDLSALLQPAEAPSVENITVTAETPDFNATADDLSALLQSEVPVAV 679
1934 SPLPSKEASRPASVAESIKDEAKS-KEESRRRESVAE-----KSPLPSK 1976
680 EENAAETIVADDLSALLQPAEAPSVENITVTAETPDFNATADDLSALLQPAEVPVAVEN 739
1977 EASRPASVAESIKDEAKSK---EESRRRESVAESK-----PLPSKEAS 2016
740 VTKTVAEIPDFNATADDLSALLQSEVPVAVENAAETITLTPDSNTSEADALPDFLKDGE 799
2017 RPASVAE---SIKDEAKSKESRRRESVAESK-----PLPSKEASRPASVAESIKDEA 2066
800 EETVDMYILSEENIPNNADTSPSESGVSDAPSEAKYDLAEMYLEIGDRDAAAEETVQKL 859
2067 EKS-----KEESRRRESAAEKSPLSKEASRPV-----SVAESVKDEADKSKESRRRESM 2115
860 LEEAEGDVLKRAQALAE 878
2116 AESGKAOSIKGDQSPLEK 2134

Search completed: December 12, 2003, 17:33:06
Job time : 31 secs

471 ---EPEAEFAVEE---PAEPEE-----PADETATEPTAE 501
696 QPAEAPAVEENVTET-VAETSDPHTAADDLSALLQPA-----EVPVAVENVTKVAIPDF 750
502 ---AEPEAVEESIEKTEVEEESAPPAARQSSPPFARRRPPQSPSPERQOTSRAH-RDI 558
751 NATADDLSALLQSEVPVAVENAAETITLTPDSNTSEADALPDFLKDGE---ETVDWSI 807
559 TSYDEDSYRAVPPRMP-----TATSFSSWSPDPKQSYTPTSP-FVSTANKYRNEYTSGSS 613
808 YLSEENIPNNADTSPSESGVSDAPSEAKYDLAEMYLEIGDRDAAAEVQKLLLEAEAGDV 867
614 YR-----PTMYTSHPDIVATGAFSALY-----STNRLIERSRST 651
868 LKRAQALAE 877
652 RERKQAMRSQ 661

RESULT 15
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL0311128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/3; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 5.9%; Score 258.5; DB 2; Length 5327;
Best Local Similarity 18.8%; Pred No. 0.0016;
Matches 173; Conservative 144; Mismatches 325; Indels 277; Gaps 33;

QY 66 QSNLDPFSGSITVTGCEPAKALLGGSVTVSEKGLTAKVHKLGDKAVTAVSSEQAVRDPV 125
1387 ETSRPESATGSKVDETEQTS-----KKSPVSRPESSEAKDK 1424
QY 126 LVFRIGAGAOVREYTAILLDPVGYSPKTKSALSDGKTHKTAFTAESQENONAKALRKTDK 185
1425 SPFASGEASR-----PESVAVSVKDEAKBSRRESIAKTHKDESSLDKAKEQESR 1475
QY 186 KDSANAAVKPAYNGKTHTVRKGETVKQIAA---AIRPKHLT-----LEQVADALLKA 234
1476 RESLAESIKP-----ESGIDKESALAKESARPSVTDKSKESRRESIAESL--- 1523
QY 235 NPNVSAHGLRAGSVLHPLNRIKAOQPKPTAKPKAETASMP-----SE 280
1524 -----KAESTKDEKAPPSPKESARPGSVVSVKDETEKSK 1559
QY 281 PSKQATV---EKP-----VEKPEAKV-----RAPEAKAKP-----AV 310
1560 PSRRSIAESAKPPIEFREVSRRPESVIDGIDKESAKPSRRDSPASKASRPESVLESV 1619
QY 311 RPEPVAANTAASETAAB-----SAPOEAAASAIPTPTDETG---N 348
1620 KQEPKSTKSRRESVAESFKADSTKDEKSLTSDISRPESAVENVMAPPKETSRRPES 1679
QY 349 AV-SEPEQVSAEETEESGLFGGSYTLILAGGAALLILLRLAQSKKARTEESVPE 407
1680 AVGSMKDESMKSPSRRESVKDGA-----AQSRSTR-PASVAE 1717

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:25:46 ; Search time 18 Seconds
(without alignments)
2299.083 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCPMPTKTDCT.....EEAEQDVLRQAALAEQLGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276.5	6.3	2468	1	MAPB_HUMAN
2	269	6.1	2459	1	MAPB_RAT
3	263.5	6.0	2464	1	MAPB_MOUSE
4	257	5.8	1337	1	DEXT_STRO
5	253	5.7	1367	1	AMVH_YEAST
6	244	5.5	1616	1	P200_MYCE
7	243	5.5	865	1	CFN_DROME
8	240	5.4	704	1	NPL1_RAT
9	239	5.4	1664	1	SLP1_CLOTM
10	236	5.4	6632	1	UN89_CABEL
11	232.5	5.3	1849	1	IGA4_HAEN
12	231.5	5.0	1018	1	HMW1_MYCPN
13	221	5.0	1087	1	NFH_MOUSE
14	220.5	5.0	831	1	NFH_RAT
15	220	5.0	1020	1	NFH_HUMAN
16	219.5	5.0	8545	1	ANCL_CABEL
17	219	5.0	1781	1	AK12_HUMAN
18	219	5.0	2004	1	MOZ_HUMAN
19	215.5	4.9	1159	1	N124_SCHPO
20	215	4.9	1694	1	IGA0_HAEN
21	214	4.9	1233	1	YF16_YEAST
22	213	4.8	1702	1	IGA2_HAEN
23	210.5	4.8	442	1	ENB_DROME
24	210.5	4.8	1238	1	SBC_C_RHOCA
25	209.5	4.8	3924	1	ANK2_HUMAN
26	209	4.7	1192	1	RTN4_HUMAN
27	207.5	4.7	1411	1	TCOF_HUMAN
28	206	4.7	5147	1	PCLO_HUMAN
29	205	4.7	1861	1	MAP2_RAT
30	202.5	4.6	1385	1	FAT1_SCHPO
31	202	4.6	705	1	ICAL_BOVIN
32	202	4.6	5085	1	PCLO_RAT
33	201.5	4.6	2090	1	N214_HUMAN

34	201.5	4.6	2805	1	MAPA_HUMAN	P78559 homo sapien
35	201.5	4.6	3381	1	PGCV_BOVIN	P81282 bos taurus
36	201	4.6	778	1	CTR6_HAEN	Q9hr92 halobacteri
37	200.5	4.6	1690	1	C190_DROME	Q9vj54 drosophila
38	200	4.5	670	1	VG50_HSV1	Q00130 ictaluriid h
39	200	4.5	2774	1	MAPA_RAT	P34926 rattus norv
40	197	4.5	1972	1	P531_HUMAN	Q12888 homo sapien
41	197	4.5	5120	1	PCLO_CHICK	Q9pu36 gallus gall
42	196.5	4.5	910	1	DNUM_MYCPN	P75354 mycoplasma
43	196.5	4.5	1125	1	MAP4_MOUSE	P27546 mus musculu
44	196.5	4.5	3256	1	KI67_HUMAN	P46013 homo sapien
45	196	4.5	5376	1	ZAN_MOUSE	O88799 mus musculu

ALIGNMENTS

RESULT 1
MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP1B) [Contains: MAP1 light chain
DE LC1].
GN MAP1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.";
RL Genomics 22:273-280(1994).
CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -!- SIMILARITY: TO MAP1A.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L06237; AAA18904.1; --
Genew; HGNC:6836; MAP1B.
DR TM; 157129;
GO: GO:0005875; C: microtubule associated complex; TAS.
DR InterPro; IPR000102; MAP1B neuraxin.
DR Pfam; PF00414; MAP1B neuraxin; 10.
DR PROSITE; PS00230; MAP1B NEURAXIN; 6.
KW Microtubules; Repeat; Phosphorylation.
CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
FT

```
FT REPEAT 1878 1894 MAPIB 1.
FT REPEAT 1895 1911 MAPIB 2.
FT REPEAT 1912 1928 MAPIB 3.
FT REPEAT 1929 1945 MAPIB 4.
FT REPEAT 1946 1962 MAPIB 5.
FT REPEAT 1963 1979 MAPIB 6.
FT REPEAT 1997 2013 MAPIB 7.
FT REPEAT 2014 2030 MAPIB 8.
FT REPEAT 2031 2047 MAPIB 9.
FT REPEAT 2048 2064 MAPIB 10.
FT DOMAIN 589 790
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match
Best Local Similarity 6.3%; Score 276.5; DB 1; Length 2468;
Matches 185; Conservative 125; Mismatches 333; Indels 277; Gaps 35;

QY 150 PKYSALSDGKT--HRKTAETAESQENONAKLRKTDKDSANAAPVAYNGKTHTVRK 207
DB 656 PKKEVAKEDKTPKKEKPKKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEV 715
QY 208 ETVKQIAAAIR-PRHLTLEQVADALLKANPNVSAHGLRAGSVLHIFNLRIKAEQPKP 266
DB 716 EEKEVKEEKEPKKEIKKLPKDAKKSSTP-----LSEAKKPA 753
QY 267 TAKPKAETASMPSPSQATV-----KPVEKPEAKVAAPAEKAEKPAVRPEVPAAN 319
DB 754 ALPKPV-----PKKEEVKDKDSVAAKPKKPKKIKVKKEGKAAE-AVAAAAGTGT 804
QY 320 TAASETAESA-----PQRAAASAIPTDPTGNAVSEPVQVSAEETSGLFGGSYT 373
DB 805 TAAVMAAGIAIAIPAKLEAERLSMSPELTKDPELKAEVDVTKDKP-----856
QY 374 LLLAGGGAALIALLLLRLAQSKARRTESVPEEPDLD-----DAADGGEIITFAVET 429
DB 857 -----QLELIEDEKLEKTEPVEAVVIQKREVTKGPAESPDEGITTEGECEC 905
QY 430 PATP-----EPAPKNDVNTLAI-----DGE-----450
DB 906 RQTPEEPFVKQGVDDIEKFEDEGAFEESETDYBEKAEETEAEPREDEGHEVCVS 965
QY 451 -----SEERLSAKOTPDV-----ETDTSNIDIDFDSLAAANGILSGALTDRET 497
DB 966 ASKHSPTDEESAKAEADAYTREKESVAGDDRAEDMDZ-ALEKG-----EAEQS 1016
QY 498 QKRADAADNALTESTDSVYEPETFPNYPVVEIVID-----TPESVVAQT 541
DB 1017 EEEADEEDKAEADAREEEYEPEKMEAEYVMAVVDKAAEAGGAEEQYGLTTPTKQLGAQS 1076
QY 542 AENKPEVDTDFSNLPSNNH-----IGTEETASAKPASPGLAGFLKA-----585
DB 1077 PGREPASSIHD--ETLPGGSESEATPASDEENRDQFEFTATSGYTQSTIEISSEPTMD 1134
QY 586 --SPETILEKTVAEQVPEELHDLKV--YETDAVAE---TAPETP-----625
DB 1135 EMSTPRDVMDETNEETESQEFVNTIKYESSLYSQEYKPADVTPLNGESEGSKTDA 1194
QY 626 ----DFNAAA-----DLSAL-----LOPA 641
DB 1195 TDGDYNASASTISPPSMEEBKFSRSLRDAYCEVKAETTLTDIKDSISAVSSEKVSFS 1254
QY 642 EAPSVENITETVATP-----DFNATADDLSALLQPSVPAVENEAEIADD-LSAL 694
DB 1255 KPSLSPPSPSPLEKTPLGERSVNFSLTPNEIKVSAEAEVAPVSPVETQVVEEHCA 1314
QY 695 LQPAEAPAEVENVETVATSTDFHTAADLSALLOPAEV---PAV-----EENV 740
DB 1315 DKTLEVSPSQSVTGSAGHTPYQSPTEKSSHL-PTVEIKPPAVPVSVFESDAKDENE 1373
QY 741 TKTVA-----EIPDFNATADDLSALLQPSVPAV--BENAAETITLETPTSNTSEADLPDF 794
DB 1374 RASVSPMBEPVFD---SESPIEKVLSPLRSPPLIGSESAYESFSLADDKASGRGAESPFE 1430
```

```
QY 795 LKDGEBETDWSIYLSSEENIPNNADTSFSESVSGDAPSEAKYDLAMYLEIGDR-----D 850
DB 1431 EKSGKQSGPDQSPVSE-----MTSTSLYQDKQEGKSTDFAPIKEDFGQKKTTDD 1480
QY 851 AAAETVQKLL---EEAEGDV 867
DB 1481 VEAMSQPALALDERKLGDV 1500

RESULT 2
MAPB RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created);
DT 16-OCT-2001 (Rel. 40, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
RN [4]
RP FUNCTION: The function of brain MAPS is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -! SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -! TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -! DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -! INDUCTION: By nerve growth factor.
```

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: LCI is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: TO MAP1A.
 CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
 CC 2459) was originally described as neuraxin in Ref.3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U52950; AAB17068.1; -
 CC EMBL; X60370; CAC16162.1; -
 CC EMBL; X16623; CAA34620.1; ALT_SEQ.
 CC PIR; A56577; A56577.
 CC InterPro; IPR00102; MAP1B neuraxin.
 CC Pfam; PF00414; MAP1B neuraxin; 10.
 CC PROSITE; PS00230; MAP1B NEURAXIN; 8.
 CC Microtubules; Repeat; Phosphorylation.
 CC CHAIN ? 2459 MAP1 LIGHT CHAIN LCI.
 CC FT REPEAT 1869 1885 MAP1B 1.
 CC FT REPEAT 1886 1902 MAP1B 2.
 CC FT REPEAT 1903 1919 MAP1B 3.
 CC FT REPEAT 1920 1936 MAP1B 4.
 CC FT REPEAT 1937 1953 MAP1B 5.
 CC FT REPEAT 1954 1970 MAP1B 6.
 CC FT REPEAT 1988 2004 MAP1B 7.
 CC FT REPEAT 2005 2021 MAP1B 8.
 CC FT REPEAT 2022 2038 MAP1B 9.
 CC FT REPEAT 2039 2055 MAP1B 10.
 CC FT DOMAIN 559 1035 GLU-RICH.
 CC FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 CC KKEE AND KKEI/V REPEATS).
 CC
 CC DOMAIN 2224 2312 LYS-RICH.
 CC FT CONFLICT 127 127 M -> V (IN REF. 1).
 CC FT CONFLICT 140 140 T -> S (IN REF. 1).
 CC FT CONFLICT 2112 2112 R -> K (IN REF. 3).
 CC FT CONFLICT 2169 2169 L -> I (IN REF. 3).
 CC SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DED8BA2 CRC64;

Query Match 6.1%; Score 269; DB 1; Length 2459;
 Best Local Similarity 19.6%; Pred. No. 0.00082;
 Matches 189; Conservative 131; Mismatches 283; Indels 362; Gaps 42;
 94 TVSEKGLTAKVHLGDKAVIA-----VSEQAVRDPVLVFRIGAGVREYTAILDPVGY 148
 600 SVTEKEVPSKEEQSPVKAEEAKATESKPKVKDKVVKVKKIKTKPEEKKE----- 651
 149 SPKTKSALSDGKTHRK--TAPTAESQENQNAKALRKTDKDSANAAVK--PAYNGKTHV 204
 652 KPKKEVAKKEDKTLKDKDKPKKEAKKEIKKEIKKEEKELKKEVKKETPLKDAKKE-V 710
 205 RKGETVKQIAAIRPKHLTLQVADALLKANPNVSAHGLRAGSVLHPLNLRITKAQPK 264
 711 KADEK-KEVKEEKEPKKEIKKIKDKIKKSTP-----LSDTKK 747
 265 PQTAKPKAETASMPSEPSKQATVE--KPVKEPEAKVAAPAEKAEKPAVRPVPVPAATAA 322
 748 PAALKPK---VAKKEEPTKKEPIAGKLKDKGVKVKIKKEGKTTAAAA--TAVGTAAVA 802
 323 SETAESAPQ---EAAASAITPTDPTGNAVSEPVQVSAEE----- 361
 803 AAGVAASGPAKELEARSIMSPED-----LTKDFEELKAEIDVAKDKIQLELIEDEE 857

QY 362 ---ETESGLFGSYTLILLAGGGAALIALLLLRLAQSKRARRTERSVPEEPDLD---D 414
 DB 858 KIKETEPG-----EAYVIQKETEVSQKSAE 882
 QY 415 AADDCGIIITFAEVEPTATP--EPAPKNDVNDTLALD-----GESEELISAKQ 459
 DB 893 SPDEGIIITTEGECEQETPEELEPVKEQGVDDIEKFDEGAGFESSEAGDYEEKAETEE 942
 QY 460 TFDVETTPSRNRIIDFLDLSLAAQNGILSGAL-----TQDEETQKRAADNMAIESTSV 514
 DB 943 AEEPEDEGDN-----VSGASKHSPTEDDEEIAK-AEADVHIKKEKRESV 985
 QY 515 -----YEPETFPYNPVEVID-- 531
 DB 986 ASGDDRAEEDMDEALEKGEAQSEEEGEEDKAEDEEDHEPDKTEADYVMAVVDKA 1045
 QY 532 -----TPEPEVAQTAENKPEVTDTDFSNLPSNNH-----IGTEETASAKP 573
 DB 1046 AEAGVTEQDYDFLGTTPAKQPGVQSPSPREPASSIHD--ETLPGSSESEATASDEENREDQP 1103
 QY 574 ASPSGLAGFLKA-----SSPETILEKTVAEVQTPDELHDFLKV--YE--- 613
 DB 1104 EEFATSGYTQSTTIEISSEPTFMDMSTPRDVTMTDTNNEETESPSQEFVNITKYESSLY 1163
 QY 614 -----TDAVAETAPETPDFAAADDL-----SAL- 637
 DB 1164 SQEYSKPVVASFNGLSGSKTDATDGRDYNASASTISPPSSMEEDKFKSALURDAYREE 1223
 QY 638 -----LQPAEAPSVBEENITETVAETP-----DFNATADDLSALLQP 673
 DB 1224 TDVKTGAELDIKVDSDERLSPAKSPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1274
 QY 674 SEVPAVEENAAEIVADDLSALLQPAEAPAVENVTETVAETSDPDHTAADDLSALLQPAEV 733
 DB 1275 NEIKASAEGEATV-----VSPGVTVQAVVEE--HCASPEEKTLL---EV 1312
 QY 734 PAVENVTKTVAEIPDFNATADDLSALLQPEVPAVEENAAEITLETDPDSTSEADALP- 792
 DB 1313 VPSQSVTGSAGHTPYQSPDTEKSSHL-PTEV-----TENQAQVPEV 1353
 QY 793 ---DFLKQEEETVDWSIYLSSEENIPNNADTSFSPSEVSGSD--AP---SEAKYDLAEWYL 844
 DB 1354 SFEP-TEAKDENERSISSPMDEVPV---DSSEPIEKVLSPRLSPPLIGSESAY---EDFL 1406
 QY 845 EIGDR 849
 DB 1407 SADDK 1411
 RESULT 3
 MAPB MOUSE
 ID MAPB MOUSE STANDARD; PRY; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
 DE [Contains: MAP1 light chain LCI].
 GN MAP1B OR MTAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A., AND DOMAIN.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RT Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
 RT and tau."
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.

[illegible]


```

Db 676 SVNDVKSLETKTTSVELNHEEIGNFINLVDSEKEVQEQPTTQLETDSEFVLPTQYIVE 735
QY 144 D-----PVGSPKTKGAL-----SDGKTRKKTAPTAEQ-----EN-----174
Db 736 DSFTESAETPNBESQKOTLFEISQTEVETESNVPTVEQTKLFEHQDENNLFTPLP 795
QY 175 -----QNAKALRTDKKOSANAAYKPAYKGTHTVRKGETVKVQIAAARPKHILTLRQ 226
Db 796 LDIETIESNALFDSKDEKSSDSLEQTF-----KEIKLDS 833
QY 227 VADALLKANPNVSAHGLRAGSVLHPIPNLRKAEQPKQTA-----KPKAETASM 277
Db 834 TVEV-----POSSQVEATFDTVQPEAVEFDEIKTQELQPEATEVVFDDHFDQVQPEQT 888
QY 278 PSEPSKQATVEKPKPEAKVAP-----EAKAEKAVAPPEPT 316
Db 889 PQEAKFDSVPEIPOSQAEFAEQISDEIKLEKTEAVFDHQLENQSEETVVTPTVT 948
QY 317 AANTAASETASAPQEAASAIIDTPTDTGNVSEPV-----EQVSAEETESGLFGG 370
Db 949 APEPETIETQLE-----PSESDQF-----SEPALDQNHPEIVIAVEVQ---IFDG 990
QY 371 SYTLALLAGGAALIALLLLRQAQSKRARTESVPEBPDLDRAADGI-----BIT 423
Db 991 TKL-----EDLKLEAFNDVNEVQPKTEAET 1021
QY 424 F---AEVETPATPEAPKPNVDNTLALDGESEELSA-----KQTFDVEDTDSNR- 471
Db 1022 FDETKLEQETSSSEPLSTELKSEATFNVSAEAVFEKPKQLTQTEKILEEBSKPEP 1081
QY 472 -----IDLDFOSLAAQNGILSGALTQD---EETQKRADADNAIESTDSVVEPTFNPY 523
Db 1082 VDQLITEASFDTV--KHEAVFDKNQQTQTEGLEEPQVSSEAEVVDDTDTTVCPEA---1135
QY 524 NPVEVITPTPEPESVAQT-----AENKPTVDTDFSDNLPNNHIGTETASAKPASGL 579
Db 1136 -----VFQV-QEKTTEVKFDDVNDQKQVISEPQVEQGEAVF---EPSAEAKFDS- 1184
QY 580 AGFLKASSPETILE--KTVAEVQ-----TPEELHDFLKVET-----DAVAETAPETPDF 627
Db 1185 VESVQDSQPEVLEVOQTELOIOPVESQPEATFTDTPQEQPEAKFDSPVETV-EQPEF 1243
QY 628 N-----AAADD-----LSALLQPAE-----APSV 647
Db 1244 SSEPTQCHVESEASDEPNYDFDEPNYDQPSYDSDLPQSEPVQVDEPNYDFEPNYE 1303
QY 648 ENITETVAETPDENATADDLSALLQPSVPAVEENAAIIVADDLSALLQPAE-APAVEEN 706
Db 1304 ---IESKPESEPFQFQVEQ-----QPGE--AVFEPSEAKFDSPVESVQDSQPEPLLEE- 1352
QY 707 VETVAETSDPHTAADDLSALLQPAEVP--AVEENVTKVTAIEIPDENATADDLSALLQPS 764
Db 1353 -VQTEIOPVESQPEATFTDTPQEQPEAKFDSPVETIOE-FQVSEPE---VVQPN 1407
QY 765 EUPAVENAAITLETDSNTSEADALPDFLDGEEETVDSWISYLSSEN 813
Db 1408 ----FEERKPTVLEEPQADBIQPEA-----SEESLDWELLVGNNS 1445

```

RESULT 7

```

CPN_DROME
ID - CPN DROME STANDARD; PRT; 865 AA.
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Calphotin.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;

```

```

[1]
SEQUENCE FROM N.A.
RP STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
[2]
SEQUENCE FROM N.A.
RP STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -! FUNCTION: Might function as a calcium-sequestering "sponge" to
regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
of Ca(2+) per mole of protein.
CC -! SUBUNIT: Homodimer (Probable).
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -! TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCELLI.
CC -! DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
DR EMBL; L02111; AAA28405.1; -.
DR EMBL; L05080; AAA28420.1; -.
DR PIR; A47282; A47282.
DR PIR; A47283; A47283.
DR FlyBase; FBgn0010218; Cpn.
DR GO; GO:0005509; F:calcium ion binding activity; IDA.
KW GO; GO:0005509; F:calcium ion binding activity; IDA.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;
Query Match 5.58; Score 243; DB 1; Length 865;
Best Local Similarity 24.28; Pred. No. 0.0027;
Matches 172; Conservative 84; Mismatches 285; Indels 170; Gaps 32;
QY 234 ANPNVSAHGLRAGSVLHPIPNLRKAEQPKQTAETASMPSEPSKQATVEKPKV 293
Db 60 ATVTIPAPADIAASVTVP-----ASVAPPVVVAATPPAASVSTFVAVAQIPVAVSA 112
QY 294 PEAKVAPEAKPKAVRPEVP--AANTAASETAAESAPQEAASAIIDTPT-----D 344
Db 113 P---VAPPVAAATPTPVQ---IPVAPVIAATPPVAAAPTAAVTPVISPIASPPVPA 166
QY 345 ETGNVSEPVQVSAEETESGLFGGSYTLLAGGGAALIALLLLRQAQSKRARTES 404
Db 167 NTTPVAAAPVAAVPAAPVAVP-----VLAPAVAPAVAPVAVPAPPVPAEIPVAT 218
QY 405 VPE-EPFDLDD-----AADDGIEITFAEVTETATPEPAKPNVDNTLALDGESEEE 454
Db 219 IPCVAPLIPESVVAATKPLAAAPVAVPAPPVAVPAPPVAVPAPPVAVPAPPVAVP 278

```


Qy	648	ENITETVAETPDFNATADLSALQPSEVPAVEENAA--EIVADDLSALLQ-----PAAE	700
Dd	444	EATKKSVTTPKARVTAKAAPSL-----PAKQAPRAGGSSDSSSSSEEKKTPPKP	497
Qy	701	PAVENVTETVAETSDFHTAADDLSALLQPAEVPAVERNVTVAEIPDNFNATADLSL	760
Dd	498	PAKKAAGAAMPKPTPVKKAASSSSSSSSDSSEEEK-----KKPKSKATPKPQA-	550
Qy	761	LQSPSEVPAVEENAA----EITLETSPDSNTSEADALP-DFLKDGEETVDWSYTLSEENIP	815
Dd	551	-KANGVPASQNGKAGKESEEEEDTEQNKAAGTKPGSGKKKHNETADEAATPOSKVK	609
Qy	816	NNATSPSPESVGSDAPEAKYDLAMYLEIGDRDAAEVTOKLLERAEQGVLKRA	871
Dd	610	LOTENTPPRRKKGKRASSPRFRVREEIEVDSR--VADNSPDKARGAAGDWGERA	663

RESULT 9

ID	SUP1_CLOTH	STANDARD;	PRT; 1664 AA.
AC	Q06852;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).		
DE	CULPB.		
GN	OcupB.		
OS	Clostridium thermocellum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1515;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NCIB 10682;		
RA	MEDLINE=93209931; PubMed=8458832;		
RX	Fujino T., Beguin P., Aubert J.-P.;		
RT	"Organization of a Clostridium thermocellum gene cluster encoding the		
RT	cellulosomal scaffolding protein Cipa and a protein possibly involved		
RT	in attachment of the cellulosome to the cell surface.";		
RL	J. Bacteriol. 175:1891-1899(1993).		
CC	-I- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.		
CC	-I- SUBCELLULAR LOCATION: Cell wall.		
CC	-I- SIMILARITY: Contains 4 S-layer homology (SLH) domains.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; X67506; CAA47841.1; -	
DR	PIR; T18262; T18262.	
DR	InterPro; IPR001119; SLH.	
DR	Pfam; PF00395; SLH; 3.	
DR	PROSITE; PS01072; SLH_DOMAIN; 2.	
KW	Cell wall; S-layer; Signal; Repeat.	
FT	SIGNAL	1..28
FT	CHAIN	29..1664
FT	DOMAIN	36..763
FT	REPEAT	36..191
FT	REPEAT	207..363
FT	REPEAT	409..565
FT	REPEAT	607..763
FT	DOMAIN	771..1377
FT	DOMAIN	1378..1449
FT	DOMAIN	1453..1494
FT	DOMAIN	1495..1565
FT	DOMAIN	1566..1625
FT	DOMAIN	1626..1646

FT	APPROXIMATE TANDEM REPEATS OF	
FT	T-P-S-D-E-P.	
FT	GLY/PRO/SER/THR-RICH.	
FT	SLH 1 (INCOMPLETE).	
FT	SLH 2.	
FT	SLH 3.	
FT	SLH 4 (INCOMPLETE).	

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31058; AAB0542.1; -;
DR EMBL; AF003131; AAB54132.2; -;
DR PDB; 1FHO; 20-DEC-00.
DR WormPep; C09D1.1; CE30426.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; ph; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH 2; 1.
DR PROSITE; PS50835; IG-LIKE; 49.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.

FT DOMAIN 2071 2171 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEY -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).

DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL1 protease).
GN	IGA.
OS	Haemophilus influenzae.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC	Pasteurellaceae; Haemophilus.
OX	NCBI_TaxID=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NHTI HK61;
RX	MEDLINE=92234949; PubMed=1373717;
RA	Foulsen K., Reinholdt J., Kilian M.;
RT	"A comparative genetic study of serologically distinct Haemophilus
RT	influenzae type 1 immunoglobulin A1 proteases.";
RL	J. Bacteriol. 174:2913-2921(1992)
CC	- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC	PRODUCING INTACT PC AND FAB FRAGMENTS.
CC	- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC	certain Pro- -Xaa bonds in the hinge region. No small molecule
CC	substrates are known.
CC	- SUBCELLULAR LOCATION: Secreted.
CC	- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC	SIMILARITY).
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	EMBL; M87491; AAA24968.1; -
DR	MEROPS; S06.001; -
DR	InterPro; IPR006315; Autotransport.
DR	InterPro; IPR005546; Autotransporter.
DR	InterPro; IPR000710; IGA_S6.
DR	InterPro; IPR004899; Pertactin.
DR	Pfam; PF03797; Autotransporter; 1.
DR	Pfam; PF02395; IGA1; 1.
DR	Pfam; PF03212; Pertactin; 1.
DR	PRINTS; PR00921; IGASERPTASE.
DR	TIGRPMs; TIGR01414; autotrans_bar1; 1.
KW	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT	SIGNAL 1 25 POTENTIAL.
FT	CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT	PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT	ACT_SITE 299 299 PROBABLE.
SQ	SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150A6A CRC64;
Query Match	5.3%; Score 232.5; DB 1; Length 1849;
Best Local Similarity	19.2%; Pred. No. 0.018;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;	
QY	35 LKNNQIKLIAASVAVASFOAHAG-----LGGINI-QSNLDEFFSGSITVT 80
DB	767 LYSGRNVANITSNITASNNAQVHIGYKTDGTVCVRSDYTGVTCHNSLSEKALNSFNPT 826
QY	81 GBEAKA-LLGGGVTSEKGLTAKVHKLGDKAVIASSEQAVROPVLVFRIGAGAQVEEY 139
DB	827 NLRGNVNLTENASFTLGRANLFGTIQSTIGTSQVNLKNSH-----WHLTGSNNVQL 878
QY	140 TAILDPVGYSPKTKSALSDBGKTHRKTAPTAESQENQAKALRK-----TDKK 186
DB	879 N-----LNGHILNAQNDANKVTYNTLTNSLSGNSFYWVDFTNNK 923
QY	187 DSANAAPVANGKTHV--RKGETVQIOIAAIPKH--LTFQVADA-----LLKANP 236
DB	924 SNKVYVNSATGNFTLQVADTGE-----PNHNELTFFDASNATRNNEVTLANG 973
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL1 protease).
GN	IGA.
OS	Haemophilus influenzae.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC	Pasteurellaceae; Haemophilus.
OX	NCBI_TaxID=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NHTI HK61;
RX	MEDLINE=92234949; PubMed=1373717;
RA	Foulsen K., Reinholdt J., Kilian M.;
RT	"A comparative genetic study of serologically distinct Haemophilus
RT	influenzae type 1 immunoglobulin A1 proteases.";
RL	J. Bacteriol. 174:2913-2921(1992)
CC	- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC	PRODUCING INTACT PC AND FAB FRAGMENTS.
CC	- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC	certain Pro- -Xaa bonds in the hinge region. No small molecule
CC	substrates are known.
CC	- SUBCELLULAR LOCATION: Secreted.
CC	- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC	SIMILARITY).
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	EMBL; M87491; AAA24968.1; -
DR	MEROPS; S06.001; -
DR	InterPro; IPR006315; Autotransport.
DR	InterPro; IPR005546; Autotransporter.
DR	InterPro; IPR000710; IGA_S6.
DR	InterPro; IPR004899; Pertactin.
DR	Pfam; PF03797; Autotransporter; 1.
DR	Pfam; PF02395; IGA1; 1.
DR	Pfam; PF03212; Pertactin; 1.
DR	PRINTS; PR00921; IGASERPTASE.
DR	TIGRPMs; TIGR01414; autotrans_bar1; 1.
KW	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT	SIGNAL 1 25 POTENTIAL.
FT	CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT	PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT	ACT_SITE 299 299 PROBABLE.
SQ	SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150A6A CRC64;
Query Match	5.3%; Score 232.5; DB 1; Length 1849;
Best Local Similarity	19.2%; Pred. No. 0.018;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;	
QY	35 LKNNQIKLIAASVAVASFOAHAG-----LGGINI-QSNLDEFFSGSITVT 80
DB	767 LYSGRNVANITSNITASNNAQVHIGYKTDGTVCVRSDYTGVTCHNSLSEKALNSFNPT 826
QY	81 GBEAKA-LLGGGVTSEKGLTAKVHKLGDKAVIASSEQAVROPVLVFRIGAGAQVEEY 139
DB	827 NLRGNVNLTENASFTLGRANLFGTIQSTIGTSQVNLKNSH-----WHLTGSNNVQL 878
QY	140 TAILDPVGYSPKTKSALSDBGKTHRKTAPTAESQENQAKALRK-----TDKK 186
DB	879 N-----LNGHILNAQNDANKVTYNTLTNSLSGNSFYWVDFTNNK 923
QY	187 DSANAAPVANGKTHV--RKGETVQIOIAAIPKH--LTF

```

QY 237 NV-----SAHGR-----RAGSVLHLPNLRKAEQPKQAK---PKA 272
Db 974 SVDRGAWKYKLRNVNGDYLYNVEVEKRNQTVDTTNTTNDIQADAPSQSNNEIARV 1033
QY 273 ETASMPSEPSKQATV--EKVEKPEAKVAAPAEKAEKPAVRPEFPVPAANTAA----- 322
Db 1034 ETPVPPAPATESAISEQETRP-AETAQPMEEITNTANSTETAPKSDTATQTNPNSE 1092
QY 323 ---SET-----AESAFOEAAAASAIIDPTDGTGNVSEPVQVS 358
Db 1093 SVSEETTEKVAENPPQBNETVAKNEQATEPTQNGEVAKEDOPTVEANTQTNEATQSEG 1152
QY 359 ABEETESGLGGSVTLTLAGGAALIALLLLRQAQSKRARRTEESVPEEPD--LDDAA 416
Db 1153 KTBETQ-----AETKSEPTESVTVSEMQPEKTVSQST 1185
QY 417 DDGTEI---TFAEVET-----PATPEAPKNDVNDTLALDGESEBELSAKQ 459
Db 1186 EDKVVVEKEEKAVETEETQKAPQVTSKEPPKQAPAEPEVPTDTNAEEAQAALQQTQPTT 1245
QY 460 TFDVETTPGNRIDLPDLSAAQNG-----ILSGALTQDEETQKRAADWNAIESTDS 513
Db 1246 VAAAEITSPNSKPAEETQOPSEKTNABEFTVPVWSENATQPTETEETAKVE---KEKTQE 1302
QY 514 VYRPETENPNYPVEIVDTPEPESVAQTAENKPEVTDTFSDNLPNNHIGTETASAKP 573
Db 1303 V--PQVASQSPKQ---EQPAKQQAQT---KQABPA--RENVLTNNVG-EPQQAQP 1351
QY 574 ASPS---GLAGFLKASSPETILEKTVAEVQTPPELHDFLKYVETDVAETAPEPDFNAA 630
Db 1352 QTQSTAVPTTGETAANSKPAKQQAQKQPEAPARENVSITVNTKEPQSQTSATVSTBQA 1411
QY 631 ADDLSALLQPAEASVEENITETVAEPDFNATADDLSALLQPEVPAVENAAEIVADD 690
Db 1412 KETSSNVQAPENSINTGSAATTMTET-----AEKSDKQFME---TVTEND 1454
QY 691 LSALLQPAEAPAVENVTETVAETSDFTHTAADDLSALLQPAEVPVAVENVTKTVAEIPDF 750
Db 1455 ---RQP-EANTVADNSVANNSESE--SKRRRRSVSQPKETSAEETTVASTO----- 1501
QY 751 NATADDLSALLQPSSE---VPAVEENAAEITILETPDSNTSEADALPDFLKQGEETVDWSI 807
Db 1502 ETTVDNSVSTPKPSRRTRRSVQVNSYE-PVELPTENAENAEVQ----- 1545
QY 808 YLSEENIPNAD-----TSFSESVGSDAPEAKYDLAEVLEIGDRDAAAEVTKVLEE 862
Db 1546 --SGNNVANSQFALRNLTSKNTNAVLNMAKAQF---VALNVG--KAVSQHISQLEMN 1597
QY 863 AEG 865
Db 1598 NEG 1600

```

RESULT 12

```

HMW1 MYCPN
ID HMW1 MYCPN STANDARD; PRT; 1018 AA.
AC Q50365; Q50348; Q50349; Q9R5R4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytadherence high molecular weight protein 1 (Cytadherence accessory
DE protein 1).
GN HMW1 OR MFN447 OR M2394.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-16.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=96257187; PubMed=8675025;
RA Dirksen L.B., Proft T., Hilbert H., Plagens H., Herrmann R.,
RA Krause D.C.;

```

```

"Sequence analysis and characterization of the hmw gene cluster of
Mycoplasma pneumoniae.";
Gene 171:19-25(1996).
[2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Hilbert H., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RN Nucleic Acids Res. 24:4420-4449(1996).
[3]
SEQUENCE OF 176-182 AND 188-198.
RP MEDLINE=92077440; PubMed=1743522;
RX Krause D.C., Lee K.K.;
RT "Juxtaposition of the genes encoding Mycoplasma pneumoniae
RT cytidine-acetylation accessory proteins HMW1 and HMW3.";
Gene 107:83-89(1991).
CC -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L38997; AAA61697.1; -.
DR EMBL; AR000038; AAB96042.1; -.
DR EMBL; Z32661; CAAB3580.1; -.
DR EMBL; Z32662; CAAB3581.1; -.
DR PIR; S73720; S73720.
DR Prosite; Q50365; -.
KW Cytadherence; Structural protein; Complete proteome.
FT CONFLICT 198 198 E -> I (IN REF. 3).
SQ SEQUENCE 1018 AA; 112214 MW; 3C3DEB273F9ABDE7 CRC64;

```

```

Query Match 5.08; Score 221.5; DB 1; Length 1018;
Best Local Similarity 20.18; Pred. No. 0.025;
Matches 185; Conservative 128; Mismatches 317; Indels 291; Gaps 45;

```

```

QY 112 VIASVSSQAVRDPVLVFRIGAGAQVREYTAILLDPVGYSPKTK----SALSDGKTHRTAP 167
Db 92 VLAGEQEQTEAPDYLOVVGNEA-----YGYDEAGEVWMSGYFEGDQWISTLP 140
QY 168 TAESQENQ-----NAKALRKTDKKOSANAAKPAAYNGKHTV 204
Db 141 QTEAEKQFGPEDNIETPTTASDFGLEADVPAEVAEPEVQPEVAAEFPVYQPEVA 200
QY 205 RK--GE--TVKQIAAIRPKHLLTLEQVADALLK-----ANENVSAHGLRAGSVL 250
Db 201 VEPVGETTATVEPOAVELQPE-VVVEPIVESQLQPPVEQVEMVQPEVAVFQLEV----- 255
QY 251 HIPNLNRIKAPQPKQAKPAETASMPSEPSKQATVE---KPVEKPEAKVAAPAEKAE- 306
Db 256 ---SLDPIGETAPILEQVEPCA-VQTQPEIPAEQSAVELOPEPPEVAVQSEMVAEAP 311
QY 307 KPAVRPEPVAANTASSETAESAPO-----FAASALDITDTETGNVAVS 351
Db 312 VTEAQQTETPTW-----ETIAETPQVVTPEVAVVAVVHQPPEVAVAEPL--PVEPAVAGVS 364
QY 352 E--PVEQVSAEETESGLFGGSYTIALLAGGAALIALLLLRQAQSKRARRTEESVPEE- 408
Db 365 ELIETEQVQPEVWVES-----TPVAEVQSEMVAEPEV 395

```

```
QY 409 --EPDLDDAADDGIEITFAEVETPATPEPA-----PKNDVNDTIALDGESEEEELSAKOT 460
DB 396 AVEPVEPQPQPEV-----QEVITTPVASVLEVPENPV-----VEVEQV 439
QY 461 FVETDTPSNRIDLDFDLSA-AAQNGILSGALTODEETQKRADA-----DWNALISTDSV 514
DB 440 VEPQETP---VEVQPEPVVETQEAERP---TQVVEPQQAPOPAVYENNLTERAPV 494
QY 515 YEPETENPNVEIVIDTPESVAQTAENKPEVD-----TPSDNLPENNHIH 564
DB 495 EQPEVI-----PVTV---ESQATA-TAEPQPAVAPVADMDYVLLHLDTVKNQP----- 539
QY 565 TEETASAPASPSGLAGLKKASSPETI---LEKTV---EVQ----- 600
DB 540 --QTAPOVPTPIKIE---VAESTPTVTSPEPTIAPPLFEILNNTSSDPLPVEVD 594
QY 601 -----TPEEL-----HDFLKVYETDAVETAPETPDFAAADLS 635
DB 595 FKHQHGAVGTHSFDDFTPEVGMESKTHCHSNSEVVRVSEPKTVPPPAVSSI--NIQ 652
QY 636 ALLQPAEAPSVEENITETVAETPDNATDLSALLQSEVPVPAVEENAEIVADDLSALL 695
DB 653 TVNRVVE-PTISTPTTPVVEAPALIEFVD-----TPPVETKEASSNVVDVQVQPV 701
QY 696 QPAEAPAEVNTETVAETSDPHTAAD-----LSALLQPAEVPVAEEN----- 739
DB 702 KPLMPVVEQRTTELQPTTEINLANSINDISIAELKQGRSNPAINFDFIFKMSYQMV 761
QY 740 VTKTVAEIPDF--NATADDLSALLQSEVPVPAVEENAEITLETDSNTSEADALPDFLKD 797
DB 762 VKKSPVQISDITNSKTDITNRFL-----IKKELQAELTRIENEQKAEFL----- 810
QY 798 GBEETVDMVSYLSENIENNADTSPSPSVSGDAPSEAKYDLAEMYLEIGRDRAAEATVQ 857
DB 811 ---NAKDLVSVCQKELL-----RSLNSDFTTAHRPSDS-----YEQLQKSGELVRNIQ 855
QY 858 KLLLEEAGDVLKRAQALAEQ 878
DB 856 KAILENESKI-KNIQITLKEQ 875

RESULT 13
NFH_MOUSE
ID NFH_MOUSE STANDARD; PRT: 1087 AA.
AC P13246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
EN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosfeld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit.";
RL Gene 68:307-314(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=3145094;
RA Sheidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
RT revealed by cDNA and genomic sequences.";
RL Brain Res. 464:217-231(1988).
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=Swiss Webster; TISSUE=Brain;
RA Carden M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -!- PWM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PWM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24496; AAA39813.1;
DR EMBL; M23349; AAA39813.1; JOINED.
DR EMBL; M24494; AAA39813.1; JOINED.
DR EMBL; M24495; AAA39813.1; JOINED.
DR EMBL; M35131; AAA39809.1; JOINED.
DR EMBL; M31012; CAA83229.1; ALT FRAME.
DR PIR; J0368; QFM5H.
DR MGD; MGI:97309; Nefh.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Repeat.
FT DOMAIN 1 97 HEAD.
FT DOMAIN 98 408 ROD.
FT DOMAIN 409 1087 TAIL.
FT DOMAIN 436 517 GLU-RICH (ACIDIC).
FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
FT DOMAIN 887 1087 GLU/LYS-RICH.
FT DOMAIN 98 129 COIL 1A.
FT DOMAIN 130 141 LINKER 1.
FT DOMAIN 142 239 COIL 1B.
FT DOMAIN 240 261 LINKER 12.
FT DOMAIN 262 283 COIL 2A.
FT DOMAIN 284 287 LINKER 2.
FT DOMAIN 288 408 COIL 2B.
FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).
FT CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).
FT CONFLICT 281 281 S -> T (IN REF. 2 AND 3).
FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).
FT CONFLICT 551 551 P -> PREAKSP (IN REF. 3).
FT CONFLICT 689 712 MISSING (IN REF. 3).
FT CONFLICT 714 714 G -> A (IN REF. 3).
FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).
FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).
SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38ED1CB9 CRC64;
```

```
Query Match 5.0%; Score 221; DB 1; Length 1087;
Best Local Similarity 22.4%; Pred.No.0.029;
Matches 176; Conservative 117; Mismatches 373; Indels 118; Gaps 34;
QY 50 VAASFQAHAGIGGLNIQSNLD-EPFGSITVTGSEAKALLGGGVTVSEKGLTAKVHKLG 108
DB 373 MAAQLREYQDL--LNVKALDIEIAAYRKLLGECECRIGFGPSPFLTE-GL-PKIPIS 428
```


RESULT 15
NFH_HUMAN
ID NFH_H

ID - NFH_HUMAN

Query Match 5.0%; Score 220; DB 1; Length 1020;
Best Local Similarity 20.4%; Pred. No. 0.029;
Matches 180; Conservative 128; Mismatches 409; Indels 16

Search completed: December 12, 2003, 17:31:28
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:28:47 ; Search time 45 seconds
(without alignments)
5046.362 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRPCPMWTKFTDCT.....EEAEGDVLKRAQALAEELGI 880

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4404	100.0	880	16	Q9JST3	Q9JST3 neisseria m
2	4273.5	97.0	875	16	Q9K147	Q9K147 neisseria m
3	4270.5	97.0	875	2	O86394	O86394 neisseria m
4	436	9.9	962	16	Q8XXX7	Q8XXX7 ralstonia s
5	381.5	8.7	919	16	Q9HZA6	Q9HZA6 pseudomonas
6	353.5	8.0	1621	16	Q9KTM5	Q9KTM5 vibrio chol
7	347	7.9	1951	16	Q8DB34	Q8DB34 vibrio vuln
8	334.5	7.6	927	2	O87015	O87015 pseudomonas
9	319.5	7.3	2768	5	Q9VC00	Q9VC00 drosophila
10	316	7.2	1822	2	Q07290	Q07290 streptococ
11	315.5	7.2	1786	5	Q9U0P0	Q9U0P0 plasmodium
12	312.5	7.1	673	16	Q8XTC4	Q8XTC4 ralstonia s
13	312	7.1	801	5	Q23635	Q23635 caenorhabdi
14	305	6.9	697	16	Q8PU23	Q8PU23 xanthomonas
15	295.5	6.7	1110	16	Q8ECR4	Q8ECR4 shewanella
16	294.5	6.7	753	5	Q9U7E7	Q9U7E7 drosophila

17	293.5	6.7	753	5	Q9VLL3	Q9VLL3 drosophila
18	286	6.5	1616	5	Q9VSJ0	Q9VSJ0 drosophila
19	282	6.4	1245	5	Q20684	Q20684 caenorhabdi
20	277	6.3	2262	5	Q9V4P4	Q9V4P4 drosophila
21	276.5	6.3	685	16	Q8P7R3	Q8P7R3 xanthomonas
22	274.5	6.2	1132	5	Q9W475	Q9W475 drosophila
23	272.5	6.2	1763	5	Q8I110	Q8I110 caenorhabdi
24	272.5	6.2	1829	5	Q22248	Q22248 caenorhabdi
25	271.5	6.2	1110	13	Q91255	Q91255 petromyzon
26	267.5	6.1	17352	5	Q95YM2	Q95YM2 procambius
27	267	6.1	10578	5	Q8ISF5	Q8ISF5 caenorhabdi
28	267	6.1	18519	5	Q8ISF6	Q8ISF6 caenorhabdi
29	267	6.1	18534	5	Q8ISF7	Q8ISF7 caenorhabdi
30	266.5	6.1	1634	5	Q9XW25	Q9XW25 caenorhabdi
31	265.5	6.0	688	2	Q9X4J3	Q9X4J3 ehrlichia c
32	264	6.0	638	5	Q24259	Q24259 drosophila
33	264	6.0	2055	2	O85472	O85472 abiotrophia
34	261.5	5.9	5412	5	Q9W596	Q9W596 drosophila
35	260.5	5.9	880	5	O17338	O17338 caenorhabdi
36	259	5.9	1764	2	Q93T34	Q93T34 haemophilus
37	258.5	5.9	5327	5	O76891	O76891 drosophila
38	257.5	5.8	885	5	Q9TXR9	Q9TXR9 caenorhabdi
39	257.5	5.8	908	5	Q8MXJ1	Q8MXJ1 caenorhabdi
40	257.5	5.8	930	5	O17339	O17339 caenorhabdi
41	255	5.8	1156	16	Q8F2J8	Q8F2J8 leptospira
42	253.5	5.8	682	16	Q9HYQ5	Q9HYQ5 pseudomonas
43	253	5.7	710	4	Q9H0J3	Q9H0J3 homo sapien
44	252.5	5.7	1852	3	Q9C2H4	Q9C2H4 neurospora
45	252	5.7	864	5	Q9VGC8	Q9VGC8 drosophila

ALIGNMENTS

RESULT 1

Q9JST3 PRELIMINARY; PRT; 880 AA.
AC Q9JST3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Neisseria-specific antigen protein, Tspa.
GN TSPA OR NWA2146
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85358.1; --
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01476; LysM; 1.
DR Complete proteome.
SQ SEQUENCE 880 AA; 93153 MW; AL701AFBE849338A CRC64;

Query Match 100.0%; Score 4404; DB 16; Length 880;
Best Local Similarity 100.0%; Pred. No. 9e-208;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGRLPRPCPMWTKFTDCTESNRIQPTHRGYLKKNRQIKLTAAASVAVASQAAGL 60
|||||

```
Db 1 MPAGRLPRRCPPMTKFTDCTRSNRIQPPTHGTYLKNRQIKLIAASVAVAAASFOAHAGL 60
Qy 61 GGLNIQSNLDEPPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSQEA 120
Db 61 GGLNIQSNLDEPPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSQEA 120
Qy 121 VRDPLVFRIGAGAQVREYTAIDLPVGYSPKTKSALSDGKTHRTKAPTAESEONAKAL 180
Db 121 VRDPLVFRIGAGAQVREYTAIDLPVGYSPKTKSALSDGKTHRTKAPTAESEONAKAL 180
Qy 181 RKTDKDSANAAVPAAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTDKDSANAAVPAAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Qy 241 HGRLRAGSVLHIPLNLRKAEQPKPQTAETASMPSEPSKQATVEKPKVPEAKVAA 300
Db 241 HGRLRAGSVLHIPLNLRKAEQPKPQTAETASMPSEPSKQATVEKPKVPEAKVAA 300
Qy 301 PEAKAEKPAVPEVPVPAANTAAETAESAPOEAAASALDPTDPTGNAVSEPEVQVSAE 360
Db 301 PEAKAEKPAVPEVPVPAANTAAETAESAPOEAAASALDPTDPTGNAVSEPEVQVSAE 360
Qy 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSKEARRTEESVPBEEPLDDAADDGI 420
Db 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSKEARRTEESVPBEEPLDDAADDGI 420
Qy 421 EITFAEVETPATPEAPKNDVNDTLALDGESEELSASQOTPDVETDPSNRIDLDPSLA 480
Db 421 EITFAEVETPATPEAPKNDVNDTLALDGESEELSASQOTPDVETDPSNRIDLDPSLA 480
Qy 481 AAQNGILSGALTQDEETOKRADADWNAIESDVSVEPETENPNPVEIVIDTPBPESVAQ 540
Db 481 AAQNGILSGALTQDEETOKRADADWNAIESDVSVEPETENPNPVEIVIDTPBPESVAQ 540
Qy 541 TAENKPTVDTFSDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
Db 541 TAENKPTVDTFSDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVQ 600
Qy 601 TPEELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEASVEENITETVAETPDF 660
Db 601 TPEELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEASVEENITETVAETPDF 660
Qy 661 NATADDLSALLQPSSEVPVPAENNAEIVADDLSALLQPAEAPAVENVTETVAETSDFHTA 720
Db 661 NATADDLSALLQPSSEVPVPAENNAEIVADDLSALLQPAEAPAVENVTETVAETSDFHTA 720
Qy 721 ADDLSALLQPAEVPVPAENVTETVAEIPDFNATADDLSALLQPSSEVPVPAENNAEITILET 780
Db 721 ADDLSALLQPAEVPVPAENVTETVAEIPDFNATADDLSALLQPSSEVPVPAENNAEITILET 780
Qy 781 PDSNTSEADALPDFLKDGEETVDSIYLSSENIENNADTSPPSSEVSGSDAPSEAKYDLA 840
Db 781 PDSNTSEADALPDFLKDGEETVDSIYLSSENIENNADTSPPSSEVSGSDAPSEAKYDLA 840
Qy 841 EMYLEIGRDAAAETVQKLEAEAGDVLKRAQALAEGLGI 880
Db 841 EMYLEIGRDAAAETVQKLEAEAGDVLKRAQALAEGLGI 880
```

RESULT 2

```
Q9K147
ID Q9K147 PRELIMINARY; PRT; 875 AA.
AC Q9K147
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE TspA protein.
GN NMB0341.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Pederson J.D., Hickey E.K.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson R.D., Dougherty B.A.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002391; AAF40784.1; -.
DR TIGR; NMB0341; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01476; LysM; 1.
KW Complete proteome.
SQ SEQUENCE 875 AA; 92488 MW; 1F921520C167D090 CRC64;
```

Query Match 97.0%; Score 4273.5; DB 16; Length 875;
Best Local Similarity 97.1%; Pred. No. 2.2e-201;
Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

```
Qy 1 MPAGRLPRRCPPMTKFTDCTRSNRIQPPTHGTYLKNRQIKLIAASVAVAAASFOAHAGL 60
Db 1 MPAGRLPRRCPPMTKFTDCTRSNRIQPPTHGTYLKNRQIKLIAASVAVAAASFOAHAGL 60
Qy 61 GGLNIQSNLDEPPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSQEA 120
Db 61 GGLNIQSNLDEPPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSQEA 120
Qy 121 VRDPLVFRIGAGAQVREYTAIDLPVGYSPKTKSALSDGKTHRTKAPTAESEONAKAL 180
Db 121 VRDPLVFRIGAGAQVREYTAIDLPVGYSPKTKSALSDGKTHRTKAPTAESEONAKAL 180
Qy 181 RKTDKDSANAAVPAAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTDKDSANAAVPAAYNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Qy 241 HGRLRAGSVLHIPLNLRKAEQPKPQTAETASMPSEPSKQATVEKPKVPEAKVAA 300
Db 241 HGRLRAGSVLHIPLNLRKAEQPKPQTAETASMPSEPSKQATVEKPKVPEAKVAA 300
Qy 301 PEAKAEKPAVPEVPVPAANTAAETAESAPOEAAASALDPTDPTGNAVSEPEVQVSAE 360
Db 301 PEAKAEKPAVPEVPVPAANTAAETAESAPOEAAASALDPTDPTGNAVSEPEVQVSAE 360
Qy 361 EETES ---GLFGGSYTLILLAGGGAALIALLLRLAQSKEARRTEESVPBEEPLDDA 416
Db 361 EETESGLFDGLFGGSYTLILLAGGGAALIALLLRLAQSKEARRTEESVPBEEPLDDA 420
Qy 417 DDGIEITFAEVETPATPEAPKNDVNDTLALDGESEELSASQOTPDVETDPSNRIDLD 476
Db 421 DDGIEITFAEVETPATPEAPKNDVNDTLALDGESEELSASQOTPDVETDPSNRIDLD 480
Qy 477 DSLAAQNGILSGALTQDEETOKRADADWNAIESDVSVEPETENPNPVEIVIDTPBP 536
Db 481 DSLAAQNGILSGALTQDEETOKRADADWNAIESDVSVEPETENPNPVEIVIDTPBP 540
Qy 537 SVAQTAENKPTVDTFSDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKT 596
Db 541 SVAQTAENKPTVDTFSDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKT 600
Qy 597 AEVQTPPELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEASVEENITETVA 656
Db 601 AEVQTPPELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEASVEENITETVA 660
Qy 657 TPDFNATADDLSALLQPSSEVPVPAENNAEIVADDLSALLQPAEAPAVENVTETVAETSD 716
Db 661 TPDFNATADDLSALLQPSSEVPVPAENNAEIVADDLSALLQPAEAPAVENVTETVAETSD 720
```

```

QY 717 FHATAADLSALLQPAEVPVAVENVTVAEIPDFNATADDLSALLQPSSEVPVAVENAAEI 776
D 721 FNATADDLSALLQPSSEVPVAVENAAETV-----ADDSALLQPAEAPVAVENAAEI 771
QY 777 TLETPDSNTSEADALPDFLKDGEETVDWMSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 836
D 772 TLETPDSNTSEADALPDFLKDGEETVDWMSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 831
QY 837 YDLAEMYLEIGDRDAAAEVTKVQLLEAEAGDVLKRAQALAEGLI 880
D 832 YDLAEMYLEIGDRDAAAEVTKVQLLEAEAGDVLKRAQALAEGLI 875

RESULT 3
086394
ID O86394 PRELIMINARY; PRT; 875 AA.
AC O86394;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE TspA protein.
GN TSPA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B:15:PI.16;
RX MEDLINE=99307215; PubMed=10377136;
RA Kizil G., Todd I., Atta M., Borriello S.P., Ait-Tahar K.,
RA Al'Aideen D.A.A.;
RT "Identification and characterisation of TspA, a major CD4+ T-cell and
RL B-cell stimulating Neisseria-specific antigen.";
DR Infect. Immun. 67:3533-3541(1999).
DR EMBL; AJ010113; CAA09002.2; -.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
SQ SEQUENCE 875 AA; 92548 MW; 1F920E217A677091 CRC64;

Query Match
Best Local Similarity 97.0%; Score 4270.5; DB 2; Length 875;
Matches 858; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 MPAGELPRCPMTKFTDCTRSNRITQPTTHRGYILKNNRQIKLIAASVAAAFQAHAGL 60
D 1 MPAGELPRCPMTKFTDCTRSNRITQPTTHRGYILKNNRQIKLIAASVAAAFQAHAGL 60
QY 61 GGLNIQSLNDEPFSGSIITVGTGEAKALGGGVTVSEKGLTAKVHKLGDKAVIASSQA 120
D 61 GGLNIQSLNDEPFSGSIITVGTGEAKALGGGVTVSEKGLTAKVHKLGDKAVIASSQA 120
QY 121 VRDPLVFRIGAGQVREYTAILDPVGSPTKSAISDCKTHRTKTAFTAESQENQAKAL 180
D 121 VRDPLVFRIGAGQVREYTAILDPVGSPTKSAISDCKTHRTKTAFTAESQENQAKAL 180
QY 181 RKTDKDSANAAVPAYNGKTHTVRKGBTVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
D 181 RKTDKDSANAAVPAYNGKTHTVRKGBTVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRURAGSVLHPIPNLRKAKQPKPOTAKPAETASMESEPSKQNTVEKPKVEKPAKVA 300
D 241 HGRURAGSVLHPIPNLRKAKQPKPOTAKPAETASMESEPSKQNTVEKPKVEKPAKVA 300
QY 301 PEAKAEKPAVPEPVPAAATAAETAESAPOEAAASAIDTPTDTGTNAVSEPEVQVSAE 360
D 301 PEAKAEKPAVPEPVPAAATAAETAESAPOEAAASAIDTPTDTGTNAVSEPEVQVSAE 360
QY 361 ETES-GLFAGGGAALIALLLLRQAQSKARRTEESVPPEEPDLDAA 416
D 361 ETESGLFDGLFAGGGAALIALLLLRQAQSKARRTEESVPPEEPDLDAA 420

```

```

QY 417 DDGIEITAEVETPATPEPAPKQVNDVTLALDGESEBELSAKQTFDVEDTTPSNRIDLF 476
D 421 DDGIEITAEVETPATPEPAPKQVNDVTLALDGESEBELSAKQTFDVEDTTPSNRIDLF 480
QY 477 DSLAAQNGILSGALTQDEETQKRADADWNALESTDSVYVETFPNPNVEIVIDTPEPE 536
D 481 DSLAAQNGILSGALTQDEETQKRADADWNALESTDSVYVETFPNPNVEIVIDTPEPE 540
QY 537 SVAQTAENKPTVDTDFSDNLPSNNHIGTETETASAKPASPSGLAGFLKASSPETILEKTV 596
D 541 SVAQTAENKPTVDTDFSDNLPSNNHIGTETETASAKPASPSGLAGFLKASSPETILEKTV 600
QY 597 AEVQTPPEELHDFLKVYETDAVAETAPETPDNNAADDLSALLQPAEAPVAVENITETVAE 656
D 601 AEVQTPPEELHDFLKVYETDAVAETAPETPDNNAADDLSALLQPAEAPVAVENITETVAE 660
QY 657 TPDFNATADDLSALLQPSSEVPVAVENAAEI VADDLSALLQPAEAPVAVENITETVAETSD 716
D 661 TPDFNATADDLSALLQPSSEVPVAVENAAET VADDLSALLQPAEAPVAVENITETVAETPD 720
QY 717 FHATAADLSALLQPAEVPVAVENVTVAEIPDFNATADDLSALLQPSSEVPVAVENAAEI 776
D 721 FNATADDLSALLQPSSEVPVAVENAAETV-----ADDSALLQPAEAPVAVENAAEI 771
QY 777 TLETPDSNTSEADALPDFLKDGEETVDWMSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 836
D 772 TLETPDSNTSEADALPDFLKDGEETVDWMSIYLSEENIPNNADTSFPSESVDGSDAPSEAK 831
QY 837 YDLAEMYLEIGDRDAAAEVTKVQLLEAEAGDVLKRAQALAEGLI 880
D 832 YDLAEMYLEIGDRDAAAEVTKVQLLEAEAGDVLKRAQALAEGLI 875

RESULT 4
08XXX7
ID O8XXX7 PRELIMINARY; PRT; 962 AA.
AC O8XXX7;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Probable transmembrane protein.
GN RSC1986 OR RS03415.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21691879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646067; CAD15688.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001854; Ribosomal_L29.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01476; LysM; 1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Complete proteome.
SQ SEQUENCE 962 AA; 99116 MW; EFDFF380852F0730 CRC64;

Query Match
Best Local Similarity 9.9%; Score 436; DB 16; Length 962;
Matches 238; Conservative 147; Mismatches 351; Indels 306; Gaps 41;

```



```

Db 464 LWLAVIGSALLALLVLLMILSRRAQKEEQAFAADTGEQEDALDLKGDFDITL 523
QY 425 AEVETPAPTEPAPKPNVNDTTLALOGESBEEL-----SAKQTFVETTPSNRIDLD 475
Db 524 DEPE--PQAAVAPQVEKTTAQTSDALGEADYIAYGRENQAELLQNAIYDEP-QRTDLR 581
QY 476 FDSLAA--AONGILSALTQDSETKRADADWNAITESTSVVETPFPNPNVPEVIDTPE 534
Db 582 LKIMEVYAEMGDREGFARQENELREIGGAQ-PQVEQLKSRV-----PAMVAVAAVA 631
QY 535 PESVAQTAENKPEITVDTPSNLPNNHIGTEETASAKP-ASPSGLAGFLKASSPETILE 593
Db 632 GLAGKLAQDELDSFLD-DLSLDSGH-----AAKPDAGQDLDAFSLDLDLGD 683
QY 594 KTVAEVQTPTE-ELHDFLKVYETDAVAETAPETPDPAADLQPAEAPSVEENITE 652
Db 684 DVQADLKSDSGALDLDLSDSLDLAASTPADKP-----VDDLDFGLDFAE 728
QY 653 TVAETPDNATADLQSEVPVAVENNAEIVADDLSALLQ---PAEAPAVEENYTE 709
Db 729 -LAETPS-QPKHDDLDGDFSLDLADP--EDKUSD---DDFLSLNDEVFAAAPADNEFTLD 781
QY 710 TVAETSDPHTAAD--DLSALLQPAEVPVAVENNVKTVAEIIPDNATADDLSALLQSEVP 767
Db 782 TEAAEPALSDPDDFLSLADEPTEPAPEKGEDESFQAQLDEVSAQLDELAS----- 833
QY 768 AVEENAAEITLETPOSTNT---SEADLPDFLKGEEETVDWSIYLSEENIPNNAOTSPPS 824
Db 834 -----NLDEPKSATPSFAEDAASVASALDGD-----ADDDDFD 866
QY 825 ESVGSDAPSEAKYDLAEVMYLEIGDRDAAAEVTKLLEAEAGDVLKRAQ 873
Db 867 LSGADEAAT--KLDLARAYIDMGDSEGARDILDEVL--AEGNDSQQA 911

RESULT 6
O9KTA5 PRELIMINARY; PRT; 1621 AA.
ID Q9KTA5
AC Q9KTA5;
RA MEDLINE=20406833; PubMed=10952301;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Hypothetical protein VC0998.
GN VC0998.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baes S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004181; AAF94159.1; -.
KW TIGR; VC0998; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1621 AA; 178073 MW; 2EF38BAC27A2485D CRC64;

```

Query Match 8.0%; Score 353.5; DB 16; Length 1621;
 Best Local Similarity 23.1%; Pred. No. 5.5e-09;
 Matches 226; Conservative 129; Mismatches 316; Indels 307; Gaps 50;
 QY 137 REYTAILOPVGVSPKTSALSDGKTHRTAPTAESENQNAKALRKTDKOSANAAPKPA 196

```

Db 24 RFFQRLLLPVAVMVTQTSTFVSAESIRLVGDCQVQPTPQ---YSENVIRSANM--EPG 78
QY 197 -YNGKTHTVRKGETVKQIAAAIRP-KHLTLBQVADALKANP-----NVSAGHRLAAGSVL 250
Db 79 RFFGPTSA---NQTLWSIASQLRPSSSVTVQOTLLAIYQLNFOAFENQNIH-TLIPGSTL 134
QY 251 HIPNLNRIKARQPKPTAKPAETASMPBSBKATVEKPVKEPE-AKVAAP--EAKAEK 307
Db 135 KVPSLAQISRNSTQDANVIMASHQAKLNQTPD---TPVRPVAPPAPPAVPATPKVEAQA 191
QY 308 PAVPEPVPAAANTAASETAESAPOEAAASAIIDTPT----- 343
Db 192 P---POVTP-----TAQEKAPTELKTPAKPSQSTDAEVMALAEKNHTLRML 237
QY 344 -----DETG--NAVSEPVQVSAEE-----ETESGLFGGSYTYLLAGG----- 380
Db 238 SQVQSEVSTLKEELGDNIRIRSEVERLEHERRAEASRLAPSALDNLNLSNGWLVALLA 297
QY 381 -----AALIALILLRLAQSKRARRTESVPEEPDLDDAADGIEITFAEVETPATPE 434
Db 298 LIPGLLTAIVVLLNRRSSAQENPTQNTITSEMP-----TAAPVTLG 341
QY 435 PAKMNDVNDTLALDG-----ESEELSAKQTF---DV-----ETDTPSN----- 470
Db 342 PEQTEIGDGLLDDDLDFSTDDKEENDAKAFSDEDDVFADLNETDLDLDFNLGQSDDL 401
QY 471 -----RIDLDFSLAAQNGILSGALTQDE-----ETOKRA-----DADWNAI 508
Db 402 FVGIDDDGDLDTEDFALNESANGI---SVNADDKALGLEEMERALNDVSEPTDNDLNSF 458
QY 509 ESTSVVPEPFPNPNVPEVIDTPEPE-----SVAQTAENKP----- 546
Db 459 DLADENQMS-----DDIEALLSGDEENELSGDKVQOSLDDLLASELDALDDEPATQ 513
QY 547 -ETVDTPSDNLPSNNHIGTEETASAKPASPSGLAG-----FLKA 585
Db 514 TETLDTLLNDELASL-----SEEDDDDFLSCAGVAGDQDLDLDFASIEQADLEQLEAKA 569
QY 586 SSPETILEKTVAEVQTP-----EELHDFLKVYETDAVETA-----PETPDFAAA 631
Db 570 IDETALLDEILAEOQAPLSESTELLDELDDFDKPENDFDAQTADLLQPEEPILDLEE 629
QY 632 DDLSALLQ-----PAEAPS--VEENITETV-----AETPDFNATADDLSA 669
Db 630 DSTQLLNEVLGEPVPEELASGLIDQNSTELLDELDDLDLDESIEATEFSVAPEKLSV 689
QY 670 -----LLQSEVPVAVENNAEIVADD-----LSALLQPAEA-----PAVENV 707
Db 690 EDGTDFELDELLEIEQHPEASLSPELATEDEFNSDTFIDDLNLSAPAKDLPLEPVLDENE 749
QY 708 TETVAETSDPHT-----AADLSALLQPAEYVA-----VEENVTKTV---AE 746
Db 750 AFAQADDFDFNPEIEGGLDLS---QPSALPANEFQTPQDEDMWFDEDDSSPTLEGNAE 806
QY 747 IPDNATADDLSALLQSEVPVAVENNAEITLE---TPDSNTSEA--DAL-PDFLKOGEE 800
Db 807 L-ELSSAEEDDL-----PEQTATNETADELLADLAAQPSQNTVDTSDDALAPDGLSQSVE 860
QY 801 ETVDWS-IYLSEEN-IPNNAD-----TSFPSPSVGSDAPSEAKYDLAEVMYLEIGDRDAAAE 854
Db 861 EPLTNDLELPEENDEPOLAEVTFSSAFDEQOQVETEIEPESEPELAAEASNDESULTALNE 920
QY 855 TVQKLLEAEAGDVLKRAQ 872
Db 921 L--DLPEYTEEDVLADVQ 936

```

RESULT 7
 Q8DB34 PRELIMINARY; PRT; 1951 AA.
 ID Q8DB34
 AC Q8DB34;
 DT 01-MAR-2003 (TremBLrel. 23, Created)

QY 425 AEVETPATPEAPKNDVNTLALDGESEBEL-----SAKQTFDVTETPSNRIDL 475
Db 524 DEPE-POVAAPQVEKTTAQTSALGEADIVYGRFNQAAELIQAIYDEP-ORTDLR 581
QY 476 FDSLAA-AQNGILSGALTQDEBTKRADMDNAIESTDSVYEPFNPVVEIIVDTPE 534
Db 582 LKLMVEYAEEMGDREGFARQENELREIGGAQ-PQVEQLKSR-----PAWAVAAV 631
QY 535 PESVAQTAENKETVDTDESDNLPNNHIGTBETASAKP-APSGLAGFLKASSPETILE 593
Db 632 GLAGAKLAQDELDSFSLD-DLSLDSGSH-----AAKPDAAQQLDADAFDLSLDDLGCD 683
QY 594 KTVAEVQVPE-ELHDFLVKYETDAVETAPEDPNAADDLSALLQPAEASVEENITE 652
Db 684 DVQADLKSDSGALDDLTLDSDLDLAASTPADRP-----VDDLDLFGLDFAE----- 728
QY 653 TVAETPDFNATADDLSALLQPSVEFAVEENAABIVADDLSALLQ-----PAEAPAVEENVTE 709
Db 729 -LAETPS-QPKHDDLGDGFSLLDAP--EDKLS--DOPLLSINDEVPAAPADNEFTLD 781
QY 710 TVAETSDHTRAD--DLGALLQPAEVAPEENVTKTVABIPFNATADDLSALLQ--PSE 765
Db 782 TEAABEPALSLPDDFSLADEPTPEAPPEKGEDSFAAQLDVSAQLDELASGLNDEPKSA 841
QY 766 VPAVEENAAEITLETDPDSNTSEADALPDF 794
Db 842 TFSFAEDRAV-----ASALDGDADDDF 864

RESULT 9
Q9VC00
ID Q9VC00 PRELIMINARY; PRT; 2768 AA.
AC Q9VC00;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE CG13648 protein.
GN CG13648.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrix J.F., Agbavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AB003750; AAF56376.1; -;
DR FlyBase; FBgn0039257; CG13648.
DR InterPro; IPR001007; VWF C.
DR Pfam; PF00093; VWC; 1.
DR SMART; SM00214; VWC; 4.
DR PROSITE; PS01208; VWF; 1.
SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684 CRC64;
Query Match 7.3%; Score 319.5; DB 5; Length 2768;
Best Local Similarity 20.2%; Pred. No. 5.2e-07;
Matches 198; Conservative 139; Mismatches 323; Indels 321; Gaps 40;
QY 101 TAKVHKLGDKAVIVSSEQAVRDPVLVFR-----GAGQVRE-YTA 141
Db 1141 TAKPNKIDEDVSEISTEIPKDVIMPTGITEQPLSHVKPDBEIQFVTSVPAQFDESTTA 1200
QY 142 ILD--PVGYSPKTSALSDBGKTHRTKTAFTAESQEN-----ONAKALRKTDKKDSANA 191
Db 1201 KVDKPIDESAEDKKPI--GESEDSKFDISEEDKKPVEESAEDKKPVEDSEEEKPLP 1258
QY 192 AVKPAYNGKTHVRKGETVKQIAAAI-----RPKHLTLEQVADALLK-----ANPNVSA 240
Db 1259 TVIPA--SEIEKESKPEDEKTEADFAAPTQPEATTTPAQIADTAKEVEDDDKLAITSAPV 1316
QY 241 HGLRAGSVLHPNLNRKIAQPKQTAKEAETASMPSESPKQATVEKPEKPAKAA 300
Db 1317 SGE-----DELKPADEKKRTETAQIP-----DAEIPASTDEPSSSTEL 1354
QY 301 PEAKAEK-----PAVRPEVPVPAANTAAS-ETAAES-----APOEAAAASAIPTP 342
Db 1355 PTVDLKKPEEDSTKGTPEAPESDKVPEVPTASITENEIEESDKFTTTPVAPPKISADETEP 1414
QY 343 TDETG--NAVSEPVQ-----VSABEETESGLFGSGY 372
Db 1415 TAEEDLVPAETFEPIESESEFEVSTKPPAVQGPPLPTLAPAQPEKKPVDATSTEADI----- 1469
QY 373 TILLAGGAALIALLLLRQAQSKARTESVPEEPDLDAAADDG-----IEITFAHV 427
Db 1470 -----STEPSARVEKASGETSESDNEIDAGASSTPPVPVSADED 1508
QY 428 ETPATPEAPKNDVNDTLA-LDGSESEELSAKQTFDV-ETDTPSNRIDLDLDFSLAAQNG 485
Db 1509 KTPSTEKTVAEADDKFTTVAPLAGDEESNLKLPQDIFEEBAPV-----A 1553
QY 486 ILSGALTODEBTKRADMDNAIESTDSVYEPFNPVYN----- 524
Db 1554 VTTRAAPSKDGEQKPEVEEKEPIEDGQKPIEDTSTPTSENEIEPESDRATTIAPSKEE 1613
QY 525 -----PVEIIVDTPEESVAQTAENKETVDTDFSNLPPNNHIGTBETAS 570
Db 1614 PSEFSTGAPTDEPAEPSTDAPEDESKESTPESEVPTTVAPAGEKIPTSTTPDEPTAT 1673
QY 571 AKRASPGLAGFLKASSPETILEKTVAEVQTPPELHDLFLKYETDAVETAPETPDFNAA 630
Db 1674 SAP-----VAKPDEDEVK-----ETSTEI-----PTDAPASS-----EEDENS 1707
QY 631 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPSVEFAVEENAAEIVADD 690
Db 1708 TDQI-----PSEVP---EKKETTAQTPPEEGDIVGATAAPTTSDEVPVQRLPEVLAE- 1758

```
QY 691 LSALLOPAEPAVEENVETVAETSD-----FHTAADLSALLOP----- 730
Db 1759 ---IPQSTETGIKQDETTAAPSIDRKPEVTVIBDEATVAPISBKDEKPTEEKPVE 1815
QY 731 ---AEVPAVEENVTKIV-----AEIPDFNATADDLSALLOPSEV-PAVEENAAEITLET 780
Db 1816 QKPTGESPESEEEKEKPIQDVSTEGP-VSTEASEAGSTESSEEVKPESTEGEVAEKPEDK 1874
QY 781 PDSNTSEA--DALPDF-----LKDBEBETVDMWSIYLSEENIPNNADTSFPSE----- 825
Db 1875 QPSSTAQAQVETIPEISTELPAQODKPT-----SEAPVDSDEDISAPSEKIPSVSG 1927
QY 826 -----SVGSDAPSEAKYDLAEM-----YLEIGRDAAAEVTKVQLLEEA 863
Db 1928 EEVEGPEVTTASPOAABEDELKTPAESPEPSSTDKVPETEQKPEDTKADETPESVTQVS 1987
QY 864 E-----GDVILKRAQA 873
Db 1988 DVATSTAPVAGGDIEKDEQA 2008

RESULT 10
Q07290 PRELIMINARY; PRT; 1822 AA.
ID Q07290
AC Q07290;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ept* protein.
EN Ept*.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1307;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1890;
RX MEDLINE=93328288; PubMed=8335363;
RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
RT "Repeats in an extracellular protein of weakly pathogenic strains of
RT Streptococcus suis type 2 are absent in pathogenic strains.";
RL Infect. Immun. 61:3318-3326(1993).
DR EMBL; X71880; CAA50714.1; -
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
SQ SEQUENCE 1822 AA; 192632 MW; 3838960C77641D7D CRC64;
```

Query Match 7.2%; Score 316; DB 2; Length 1822;
Best Local Similarity 19.9%; Pred. NO. 4.4e-07;
Matches 204; Conservative 165; Mismatches 363; Indels 294; Gaps 43;

```
QY 37 NNRIQKLAASV-----AYAAEQAHAGLGGNIQSNLD--BPFSGSIIVTGEAKAL 87
Db 682 DNERLKGLPDSAFVNSDGTVSDVSYA-----GGVNDGATDIKNAATNLADTRNQAKAE 737
QY 88 LGGSVTVSEKGLTAK----VHKLGDKAVIASVSEQAVRDPVLVPRIGAGAQVREYTAIL 143
Db 738 I-DTKLAEHKAIEAKRDEAFSKIDDDISLRAEQFQAADAVAA---AAGDALKE----L 789
QY 144 DPVGVSPK-----TKSALSDGKTHRTAPTASQENQONAKALRKTDKDSANAAV--- 193
Db 790 DNKATEAKEKIDKATTASEINDAKTNGEI--NLDSAAVGEKAINQAKELAKAEVENEK 847
QY 194 -----KPAV-----NGKTHTVRK---GETVVKQIAAAIRPKHLT----- 223
Db 848 AFEALEKVNNNLLEEKATVDDIKESKEVAVKINNAENTAEITAAIDEAEIAYNED 907
QY 224 -----LEQVAD-----ALLKANPNVSAHGRLRAGSVLHPIPLNLRIRKAEQPKQTA 268
Db 908 VINRAQLDALNKLKXDSSETKAIDAENPNLTPEEKAKA-----IAKVEELVNNABSILS 962
QY 269 KPAETASMPSEPSQATVEKPKVEKPAKVAAPAEK---AEKPAVRPEVPAANTAASET 325
```

RESULT 11

```
Q9U0P0 PRELIMINARY; PRT; 1786 AA.
ID Q9U0P0
AC Q9U0P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
RT protection in chimpanzees.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007010; CAB65343.1; -
DR InterPro; IPR001313; Pumilio/Puf.
```

[illegible]

QY 615 DAVETAPETPDNFNAADLSALLOPAPAPSVENITETVAETPOFNATADD-----LSA 669
 Db 602 NADAAMPVAPAPPAEAD-AAAEPAEVP-----EAPAPATADDLPHSLLEL 649
 QY 670 LIQPSVPAVEENAAEIVADDSAL 694
 Db 650 LAAPQDPATSAKVPRL--PDLELL 672

RESULT 13
 Q23635 PRELIMINARY; PRT; 801 AA.
 AC Q23635;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE ZK84.1 protein.
 GN ZK84.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Bonfield J., Burton J., Durbin R., Favello A., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Hillier L., Jier M., Johnston L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Laister N., Latreille P.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Mortimore B., O'Callaghan M.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX Kirsten J.;
 RT "the sequence of C. elegans cosmid ZK84.";
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U23181; AAC48204.1; --
 DR WormPep; ZK84.1; CE02924.
 SQ SEQUENCE 801 AA; 77123 MW; 070D8F085A71EF28 CRC64;

Query Match 7.1%; Score 312; DB 5; Length 801;
 Best Local Similarity 24.7%; Pred. No. 2.3e-07;
 Matches 205; Conservative 77; Mismatches 365; Indels 184; Gaps 33;

QY 97 EKGLTAKVHLGLKAVIAVSSEQVRDPLVFRIGAGAQVR-----EYTAILDPVGYSPKT 152
 Db 77 ERDNAAAAXKSTDAVNSASVDTGVGD-----GAVADVAPGPEPTPATAPGAAPV 129
 QY 153 KSAL-SQCKTHRKTHPTAESQENQNAKALRTDKDSANAAVKPAYNGKTHTVRKEIVK 211
 Db 130 PTAVEGSGSVGSGIP-----DEVATTVAADSGSDAPASDNSVTTTLTATEPAI 179
 QY 212 QIAAAIRPKHLTLEQVADALLKANPNVSAHGRSLRAGSVLHPLNLRKAEOPKQOTAKPK 271
 Db 180 ILPPPAEPKKI-----IPPSGVTVTSEQQAGD-----APSPAPVVEPT 220
 QY 272 AETASMPSEPSQATVEKPVKEPAKV-----AAPEAKAEKPAVRPEVPAA-NTAASETA 326
 Db 221 APTPAETPAPATEASNAVGPPEGVVDGTANAASAVAEPAVETPAPAPAAEETPAPATS 280
 QY 327 AESAPQEAASAIPTDPT-----GNAVSEPV-----EQVSAEEETESGLF 368
 Db 281 SEAP-APAPTAEETPAPETVSAAPPAANSYDSAGGDAATAPAPSSAEADAAPTDSA-- 337

QY 369 GGSYLLLAGGAALIALLLRLAQSKKARETESVPEEP-----DLDDAADDGIRI 422
 Db 338 -----ASADITTAALVDTSSSEHAAESTEAPATDIAATETTPAPSPVAPVADRA----- 385
 QY 423 TFAEVETPAT-PEPAPKNDVNDTLALDGESEBELSAKQTFDVTETDTPPSNRILDLDFSLAA 481
 Db 386 --AGYDPSPSIPEETPAPAAEDTPAPASAAAEETPAPAAAEETPAPET-----AS 434
 QY 482 AQNGILSGALTQD-----ETQKRADADMAISTDSVYEPETPNPNFVPIVIDTPE 534
 Db 435 AAPDAAGGAAPADVAAPADVATTAPETSSAQAGSYDVPSEPAVEVTAPIVESATEAPS 494
 QY 535 PESV---AQTAENKPTVDTDFSDNLPNNHICTETASAKPASPS---GLAGFLKASS 587
 Db 495 DSAAPIGPAASEPAPAPIEAPATD-----AATLETAPAPAAEPAPAAEAAGY---DA 544
 QY 588 PETILEKTVAEVTPEELHDFLKVVETDAVETAPETPOFNAAADLSALLOPAAEPSVE 647
 Db 545 PSVPEETPAPAAD-----ETPAPAAEETPAPAAEE-----TPAPAAE 590
 QY 648 ENITETVA--ETPDFNATADDL-----SALLQPSVPAVEENAA-----EIVADDLSAL 694
 Db 591 ETPATAPAAEETPAPAPAADETPAPAPAAEETPAPAPAAEETPAPAPAAEETPAPAPAAE 650
 QY 695 LQPAEAPAVEENVETVAETSDPHTAADDLSALLO-----PAEVPAAVEENVTKVAEIPD 749
 Db 651 ETPAPAPAAEYAAPVAEETPAPAPAAEETPAPAPAAEETPAPAPAAEETPAPAPAAE 706
 QY 750 FNATADDLSALLOPSEVPAAVEENAAEITL--ETPDNNTSEADALPDLKDGEBETVDWSI 807
 Db 707 AET-----PAPAPAAEAP 748
 QY 808 YLSEENIPNNAATFFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEVTVOK 858
 Db 749 YDAAQSDTDTVASSAPATS--SDASGATNYEGQOFAIQ-----AARKKVR 792

RESULT 14
 Q8PJ23 PRELIMINARY; PRT; 697 AA.
 ID Q8PJ23
 AC Q8PJ23;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE FimV protein.
 DE XAC2732.
 GN XAC2732.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2202145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Caramotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorfy H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Moreira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463 (2002).
 DB EMBL; AE011911; AM37567.1; --

Db 747 LETHLDTEDKAEEDLLKIGIAHAIDGTDTFELSDTIADRRDDANEATDEPTFTGF 806
Qy 829 SDAPSEA-KYDLAEMYLEIGDRDAAAETVQK 858
Db 807 VD--NEANNVELATATLTAATAAAAAAKVKE 835

Search completed: December 12, 2003, 17:32:27
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:31:32 ; Search time 21 Seconds
(without alignments)
2260.570 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGLPRRCPMWTKFTDCT.....ERAEGVLKRAQALQELGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 285091 seqs, 53945424 residues

Total number of hits satisfying chosen parameters: 285091

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	5.5	1616	5	US-09-820-843B-16
2	232.5	5.3	1848	6	US-10-687-046-6
3	230	5.2	1072	7	US-60-485-101-741
4	219.5	5.0	1033	6	US-10-258-899A-3455
5	219	5.0	1781	6	US-10-428-487-16
6	219	5.0	2004	7	US-60-490-890-1750
7	218	5.0	1026	7	US-60-485-101-597
8	215	4.9	1026	6	US-10-258-899A-1487
9	214	4.9	748	6	US-10-425-114A-72422
10	213	4.8	1702	6	US-10-687-046-5
11	213	4.8	1881	1	PCT-US03-27401-316
12	213	4.8	1881	6	US-10-472-928-1236
13	210	4.8	1270	1	PCT-US03-24084-4
14	210	4.8	1311	1	PCT-US03-24084-5
15	209.5	4.8	3897	1	PCT-US03-28227-4168
16	209.5	4.8	3924	1	PCT-US03-28227-4168
17	209.5	4.8	3955	1	PCT-US03-28227-4164
18	209.5	4.8	3985	1	PCT-US03-28227-4163
19	209.5	4.8	4072	1	PCT-US03-28227-4161
20	209.5	4.8	4072	1	PCT-US03-28227-4162
21	209.5	4.8	4082	1	PCT-US03-28227-4160
22	207	4.7	777	7	US-60-495-114-1620
23	206.5	4.7	778	7	US-60-495-114-1623
24	206.5	4.7	2130	7	US-60-490-890-2346
25	203.5	4.6	2828	6	US-10-454-351-21
26	202.5	4.6	732	7	US-60-495-114-1629

RESULT 1

US-09-820-843B-16

; Sequence 16, Application US/09820843B

; GENERAL INFORMATION:

; APPLICANT: Council of Scientific and Industrial Research

; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEINS USEFUL AS ANTI-INFECTIVES

; FILE REFERENCE: Q63915

; CURRENT APPLICATION NUMBER: US/09/820.843B

; CURRENT FILING DATE: 2001-03-30

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 16

; LENGTH: 1616

; TYPE: PRT

; ORGANISM: M. genitalium

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: cytoadherence-accessory protein

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: gi|1046097

US-09-820-843B-16

Query Match 5.5%; Score 244; DB 5; Length 1616;

Best Local Similarity 20.0%; Pred. No. 0.002;

Matches 178; Conservative 124; Mismatches 301; Indels 286; Gaps 43;

Qy 92 SVTVSEKGLTAKV-----HKLGDKAV-JAVSEQAVRDPVLVFRICAGAVREYTAIL 143

Db 676 SVNDVDSKLETKTSVLEHNEEIGNFINLDVSEKVEQEQPTQLETDSEFVLTYQIVE 735

Qy 144 D-----PVGSPKTKSAL-----SDGKTHRKTAFTAESQ-----EN----- 174

Db 736 DSFTESAETNEFSSEQDLEFISQEQVETSESNVPTVEQETKLFHQDNNLFTPLP 795

Qy 175 -----QNAKALRKTDKDSANAQVPAVNGKTHTVRKGETVKQIAAIRPKHILTEQ 226

Db 796 LDUTEIESNALFDSKPEKESDSELQPTF-----KEIKLDS 833

Qy 227 VADALLKANPNVSAHGRLRAGSVLHPIFNRIKAEQPKQTA-----KPKAEATSM 277

Db 834 TVEV-----PQESSQVEATDTVQPEAVFDEIKTQEQPEATTEWFDHFDPVQPEQT 888

Qy 278 PSBPSQATVEKVEKPEAKVAAP-----EAKAEKPAVRPEVP 316

Db 889 PQEAKFDSPVIEIQESSQAQFAEQISDEIKLEKTEAVFDHQLEQSQSEETVVTFTVT 948

Qy 317 AANTAASETAESAPOEAASAITDPTDTGNVSEFV-----EQVSEETEESGLFGG 370

ALIGNMENTS

Sequence 1625, Ap
Sequence 1630, Ap
Sequence 2344, Ap
Sequence 1624, Ap
Sequence 1621, Ap
Sequence 67803, A
Sequence 63112, A
Sequence 836, App
Sequence 1659, Ap
Sequence 2633, Ap
Sequence 1660, Ap
Sequence 9, Appli
Sequence 9, Appli
Sequence 837, App
Sequence 838, App
Sequence 62498, A
Sequence 310, App

27 201.5 4.6 750 7 US-60-495-114-1625
28 201.5 4.6 750 7 US-60-495-114-1630
29 201.5 4.6 2090 7 US-60-490-890-2344
30 201 4.6 791 7 US-60-495-114-1624
31 200.5 4.6 726 7 US-60-495-114-1621
32 200.5 4.6 747 6 US-10-425-114A-67803
33 198 4.5 1989 6 US-10-425-114A-63112
34 196.5 4.5 2801 6 US-10-719-993-836
35 196.5 4.5 2896 6 US-10-719-993-835
36 196.5 4.5 2896 7 US-60-485-450-1659
37 196.5 4.5 2896 7 US-60-490-890-2633
38 196.5 4.5 3061 7 US-60-485-450-1660
39 196.5 4.5 3236 1 PCT-US02-23749-3
40 196.5 4.5 3256 1 PCT-US03-35115-9
41 196.5 4.5 3256 6 US-10-701-490-9
42 196.5 4.5 3256 6 US-10-719-993-837
43 196.5 4.5 3256 6 US-10-719-993-838
44 196 4.5 453 6 US-10-425-114A-62498
45 196 4.5 10431 6 US-10-475-117-310

Db 949 AFPPETIETOLE-----PSSDOP-----SEPALDQNHPEIVTAEVQ-----IFDG 990
 Qy 371 SYTILLAGGAALIALLLRLAQSRRARTEESVPEEPDLDAAADGI-----EIT 423
 Db 991 TKL-----EDLKEANFONVENNEVQPKETEAREIT 1021
 Qy 424 F---AEVETPATPEPAPKNDVNDTLALDGESEBELSA-----KQTFDVEITTPSNR- 471
 Db 1022 FDETKELQETSSEPLSBEELKSEATFDNVSEASEAVFEKPOLEQTOKILEEPEKSEP 1081
 Qy 472 -----IDLDPDLAAAGNILLGALTOD---EETQKADADWNAIESTDSVYEPETENPY 523
 Db 1082 VDQLITEASFDIV--KHEAVFDKNQQTQEGLEEQVSEAEVVDQTTDIVGEPEA--- 1135
 Qy 524 NPVEIVIDTPEPESVAQT---AENKPEVTDVDFSDNLPSNNHIGTRETASAKPASGL 579
 Db 1136 ---VFDV-QPEKTTVEKFDVQKQVISEPQVEQPGEAFF--BPSAEAKFDPSP--- 1184
 Qy 580 AGFLKASSPETILE--KTVAEVO---TPRELHDFLKYET-----DAVAETAPETPDF 627
 Db 1185 VESVQDSQPEPVLSEVQPEIQVQESQPEATFTDTPQEQPQAKFDSPVETV-EQPEF 1243
 Qy 628 N-----AAADD-----LSALLQPAE-----APSV 647
 Db 1244 SSETQOHVESEASFDEPNYDFDQPSYDSLDLPSEPOVDYDDEPNYDEPNYE 1303
 Qy 648 ENITETVAETPDFNATADDLSALLQPEVPAVENAAEIVADDLSALLQPAE-APAVEEN 706
 Db 1304 ---IESKSEFQEPQVEQ-----QPCF--AVFEPSEAKFDSPVESVQDSQPEFLLEE- 1352
 Qy 707 VTEIVAEISDHTAADDLSALLQPAEVP--AVEENVTKTVAEIPDFNATADDLSALLQPS 764
 Db 1353 -VQTPQEPQVQESQPEATFTDTPQEQPQAKFDSPVETIQE-PQVSSEPE---VVQPN 1407
 Qy 765 EVPAVENAAETLTETPDNSNTSEADALPDFLKGDEETVDSIYLSSEN 813
 Db 1408 ---FEERKPEVILEEPOADRIQPEA-----SEESLDEWELLVGNNS 1445

RESULT 2

US-10-687-046-6
 ; Sequence 6, Application US/10687046
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme, Joseph W.
 ; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
 ; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
 ; CURRENT APPLICATION NUMBER: US/10/687,046
 ; CURRENT FILING DATE: 2003-10-15
 ; PRIOR APPLICATION NUMBER: US/10/080,505
 ; PRIOR FILING DATE: 2002-02-22
 ; PRIOR APPLICATION NUMBER: US 08/296,791
 ; PRIOR FILING DATE: 1994-10-25
 ; PRIOR APPLICATION NUMBER: US 09/839,996
 ; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 1848
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-687-046-6

Query Match 5.3%; Score 232.5; DB 6; Length 1848;
 Best Local Similarity 19.2%; Pred. No. 0.0072;
 Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;
 Qy 35 LKNNRQIKLIAASVAAVAFQAHAG-----LGGLNI-QSNIIDPEPFGSITVT 80
 Db 766 LVSGRNVANITSNITASNAQVHIGYKIGDTVCVRSYDYGTYTCHNSNLSEKALNSFNP 825
 Qy 81 GEEAKA-LLGGGSVTVSEKGLTAKVHKGDKAVIAVSSQAVRDVPLVFRICAGAQVREY 139

Db 826 NLKGNVNLTENASFTLGKANLFGTIOISIGTSQVNLKNSH-----WHLTGNSVNL 877
 Qy 140 TAILDPVGYSPKTKSALSOGKTHRTAPTAESENQAKALRK-----TDKK 186
 Db 878 N-----LTNGHILHNAQNDANKVTTNTLTVNSLSGNSFYVWDFNNK 922
 Qy 187 DSANAAVKPAYNGKTHTV--RKGETVKQIAAAIRPKH--LTLQVADA-----LLKANP 236
 Db 923 SNKVVVNSKATGFTLQVADKTCG-----PNEHNLTLFDASNATRNLEVTLANG 972
 Qy 237 NV-----SAHGL-----PAGSVLHHPNLRKAEOKPKQAK--PKA 272
 Db 973 SVDRGAWKYKLRNVGRYDLYNPEVEKRNQTVDTTNTTNDQADAPSAQSNNEETARV 1032
 Qy 273 ETASMPSEPKOATV--EKPVEKPEAKVAAPAEKAEKPAVRPEPVPAAATAA----- 322
 Db 1033 ETPVPPAPATESIAIEQEPETRP-AETAQPAEMEETNTANSTETAPKSDTATQENPSE 1091
 Qy 323 ---SET-----AESAPOEAAAASALDTPDETGNVSEPVQVS 358
 Db 1092 SVFSETTEKVAENPQENETVAKNEQEAETPTPQNGEVAKEDQPTVEANTQTNATQSE 1151
 Qy 359 ABEETESGLFGGSYTLLLAGGGAALIALLLRLAQSRRARTEESVPEEPD--LDAA 416
 Db 1152 KTEETQT-----AETKSEPTESVTVSENGPEKTVSQST 1184
 Qy 417 DDGIEI---TFAEVET-----PATPEAPKNDVNDTLALDGESEBELSAKQ 459
 Db 1185 EDKVVVEKEEKAKVETETQKAPQVTSKEPPKQAEPAPEEVPVTDNAAEQALQOQTPT 1244
 Qy 460 TFDVETDTPSNRIDLDFSLAAQNG-----ILSGALTQDEETQKADADWNAIESTDS 513
 Db 1245 VAAAEETSPNSKPAEETQOPSEKTNABPVPVWSENATQPTETEETAKVE---KEKTOE 1301
 Qy 514 VYETETPNPNPVEIVIDTPEPESVAQTAENKPEVTDVDFSDNLPSNNHIGTETASAKP 573
 Db 1302 V--EQVASQESPKQ---EQPAKPAQAT---KQAEPA--RENVITTKNVG-EQOPQAP 1350
 Qy 574 ASPS---GLAGFLKASSPETILEKTVAEVOVTPPEELHDFLKYETDAVAETAPETPDFNAA 630
 Db 1351 QTSATVPTTGETAANSKPAKPAQAKPQTEPAEENVSTVNTKEPQSQTSATVSTEQPA 1410
 Qy 631 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPEVPAVENAAEIVADD 690
 Db 1411 KETSSNVQEPAPENSINTGSATMTET-----AEKSDKPQME---TVTEND 1453
 Qy 691 LSALLQPAEAPAVENVTVETVAETSDPHTAADDLSALLQPEVPAVENVTKTVAEIPDF 750
 Db 1454 ---RQP-EANTVADNSVANNSESE--SKGRRRRSVSQPKETSAAETTVASTQ----- 1500
 Qy 751 NATADDLSALLQPSF---VPAVENAAETITLETDPDNTSEADALPDFLKGDEETVDSI 807
 Db 1501 ETTVDNSVSTPKPRSRRTFRSVQVNSYE-PVELPTENAENAENVQ----- 1544
 Qy 808 YLSEENIPNNAD-----TSFPSESVGSDAPSEAKYDLAEVLEIGDRDAAAEVTVKLEE 862
 Db 1545 ---SGNNVANSOPALNLTSKNTNAVISNAMAQAF---VALNVG--KAVSOHISOLEMN 1596
 Qy 863 AEG 865
 Db 1597 NEG 1599

RESULT 3

US-60-485-101-741
 ; Sequence 741, Application US/60485101
 ; GENERAL INFORMATION:
 ; APPLICANT: Purdue Pharma L.P.
 ; TITLE OF INVENTION: GENES ASSOCIATED WITH RESPONSES TO NEUROPATHIC PAIN
 ; FILE REFERENCE: 2755/0M584
 ; CURRENT APPLICATION NUMBER: US/60/485,101
 ; CURRENT FILING DATE: 2003-07-03
 ; NUMBER OF SEQ ID NOS: 868

QY 144 DPVGYSPKTKSALSD

343 STXDSLQRSELED---RHQADIASYQEAIOQLDDELNTKWEMAAQLREY
:

[illegible]

RESULT 6

Qy		193	VKPAYNGKTHTVRKGETVKVIOAAAIRPKHILTEQVADALLKANPNWSAHGRLRAGSVLHI	255
Dq		397	VKMALDIETIAAYFKLLLEGECRGFGIPSLPEG----	441
Qy		253	PNLNRIKAEPKPOTAKPKEATASMPSEPSKOATVEKPVKPEAKVAAPAEKAKPAVRP	312
Dq		442	---IKVKEBEKIIVKESEKETVIVEQTETEVTQVTEEEDKEAKEEGEEGBEE	498
Qy		313	EPVPAANTASETAASPOEAAAASALDTPDETGNVSE-----PVEVSJE-BETESGL	367
Dq		499	E-----AEGGE BETKSPPAEEAAS-----PEKEAKSPVKEEAKSPAEEKSPEKEAKSP-	547
Qy		368	FGGSYTULLAGGGAALTALLLLLRLAOSKRARRTEESVPER--EPDLDDAADDOIGIEITF	424
Dq		548	-----AEVKSPEKAKSPAKEBAKSPPEAKSPPEKDGKQNFO	582
Qy		425	AEVETPATPPAPKNVDNTLALDGESEELSASAKOTPDVETDTPS----NRIDLDFDSLAA	481
Dq		583	AEVKSPPKASPAAKEEAAS-----PAEAKSPKAKSPVKEEAKSPAEEKSPVKEEAKSPA	638
Qy		482	AQGILSGALTODE-ETQKRADADWNATESTDVVYPERTFNP---YNPVEIVIDTPPEPS	537
Dq		639	VKGPEKASPTKEBAKSPKAKSP-----EKAKSPEKEEAKSPEKAKSPVKAEEKSPKAK	694
Qy		538	VAQTAENK-PETVDTDFSDNLPNNNHIGITEETASAKPASPSGLAGFLK--ASSPTILEK	594
Dq		695	SPVKABEAKSPEKAKSPVKEEAKSPEKAKSPVKEEAK--SPBKAKSPVKEEAKTTPKAKSP	752
Qy		595	TVAEVQTPBELHDPLKVYETIDAVAEATPETFNFNAADDLSALLQPAEAPSVEENITETV	654
Dq		753	VKEAKSPEKAKSPEKAKTLNVKSPA-KTFAKEARSPAKFPPEKAKSPVKEEVKSPEK	811
Qy		655	AETPDFNATADDLSALLQPSVPAVEENAABIIVADDLSALLQPAEAP-AVEENVTTVAE	713
Dq		812	AKSP---LKEDAKA--PEKEIPKKEEVKS PVKBEKKPQEVVKPPEKPAEAKCAPATPK	865
Qy		714	TSDPHHTAADLSALLQPAEVPVAVBENVTKTV-----AEIPDNATADDLSALLQOP-SEVP	767
Dq		866	TEEKDKSKKE-EAPKGPAPKPVKEEBKPAVEKPKESVKEAKKEBAEDCKKYPTPEKEAP	924
Qy		768	A---VERNA-----AEITLET PDSNTSEADALPDFLDGCEERTVDWISYLSE--ENIPNN	817
Dq		925	AKVEVKEDAKPKEKTEVAKEBPDDAKAPEKSPKAEKKEAAPPEKKOTKEKAKKPEKPKT	984
Qy		818	ADTSFPSESVGSDAPSEAKEYDLAEMYLEIGDRDA	851
bq		985	EAKAKEDDKTUSKGPSPKPAEKAKSSSTDDOKDS	1018

```

RESULT 5
US-10-428-487-16
; Sequence 16, Application US/10428487
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA K.
; APPLICANT: GERBER, HANS-PETER
; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
; FILE REFERENCE: 09800080-0103
; CURRENT APPLICATION NUMBER: US/10/428,487
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 09/815,153
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,201
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-428-487-16

```

Query Match 5.0%; Score 219; DB 6; Length 1781;

```
US-60-490-890-1750
; Sequence 1750, Application US/60490890
; GENERAL INFORMATION:
; APPLICANT: Li, Martha
; APPLICANT: Rupnow, Brent A.
; APPLICANT: Webster, Kevin R.
; APPLICANT: Jackson, Donald
; APPLICANT: Wong, Tai W.
; TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
; FILE REFERENCE: D0310 BSP
; CURRENT APPLICATION NUMBER: US/60/490, 890
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 2779
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1750
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-490-890-1750

Query Match      5.0%; Score 219; DB 7; Length 2004;
Best Local Similarity 20.1%; Pred. No. 0.03;
Matches 174; Conservative 124; Mismatches 332; Indels 236; Gaps 37;

QY 95 VSEKGLTAKVHKL-----GDKAVIASSEQAVRDFVLFRICAGAGVREYTAIDPVG 148
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 ICPQDITSTLHLRLMLDFRSDQFVI--IRREKLIQDHMAKIQLN-----LRPVDV 769

QY 149 SPK-----TKSALSDGKTHKTAFTAESQBNQAKALRTDKKDSANAAVKPAYNGKTHVT 204
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 770 DPECLRWTFVINSVSVSEEEAEAGEENEPQC--QERELRISVSKSVSHENKQDSYS 828

QY 205 RKGETVQIAAAIRPKHLLTLEQVADALLKANPNVSAHGRLAGSVLHIPNLRKAKQPK 264
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 829 VESEKKPEVMAVPSVTRLSKQVLPDHSLPANQPSRRGW-----GRNKRKTQERFG 880

QY 265 POTAK--PKAETASMPSE-----PSKQATVEKPEKPEAKVAAPAKAE--KP----- 308
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 881 DKDSKLLLEETSSAQEQYGEQEKSEATQEQYTESERQLVASEEQPDGKPDLPKRL 940

QY 309 --AVRP-----EPVPAANTASETAESAPOQAAASAIPTPTDETGNVAVSEPVQVSA 361
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 941 SEGVEPWRGOLKKSPEALKCRLTGSERLPRYS-----EGDRAVLGRFSSSEEE 991

QY 362 ETESGLFGSGVTLILAGGAALIALLLLRLAQSK-----RARRTEESVPEEPDLDDAAD 417
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 992 EEPESPRSSPILTKP-----TLKRKPFLLHRRRRVRKRKHNSVVVTETIS 1039

QY 418 DGIELTFAEVETPAPPEPAPKNDVNDTIALDGESEEE-----LSAKQTFDVE 464
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 ETTEVLDPFFEDSDSERMPR--LEPTFEIDEEBEEEDENELFPREYFRLSSQDVLRCQ 1097

QY 465 TDTFSNRIDLDFDSIAAQNGLSALTQDEETOKRADADNNAIESTDSVYEPETFPYN 524
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 SSSKRSKSD-----EEDEESDDAD----- 1117

QY 525 PVEIVIDTP--EPESVAQTAENKPTVDTDFSDNLPSNNHIGTBETASAKP-----ASPSG 578
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1118 -----DTPILKPSVILRKRVKNSPLEPDTIS--TPLKKKGWPKGSRKPIHKKRPGR 1169

QY 579 LAGF-----LKASPETILEKTV-----EVQTPPEEL-----HDFLKVVET 614
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1170 KPGFKLSREIMVSTQACVIEPIVSIKAGRKPKIQESEEIVPEKEDMPLPEERKEEEM 1229

QY 615 DAVETAPETPDFNAADDLSALLQPAEAPAVENITETVAETPDFFNATADDLSALLQPS 674
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1230 QARAEAEPEGEEDAAASV-----PAASPADSSNSPETETKEPEVE-----EEE 1274

QY 675 EVPAVEENAAEIVADDLSALLQPAEAPAVENVTETVAET--SDPHTAADDLSALLQPAE 732
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1275 EKPRVSEQRQSEEQ-----QELLEEPEEE--EDAAAEATQNDHDDADDEDDHGLESTK 1328

QY 733 VPAVEENVTKTVABIPDFNATADDLSALLQPSVPAVEENAAETITLETPTSNTSEADALP 792
```

```
US-60-485-101-597
; Sequence 597, Application US/60485101
; GENERAL INFORMATION:
; APPLICANT: Purdue Pharma L.P.
; TITLE OF INVENTION: GENES ASSOCIATED WITH RESPONSES TO NEUROPATHIC PAIN
; FILE REFERENCE: 2755/0M584
; CURRENT APPLICATION NUMBER: US/60/485,101
; CURRENT FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 868
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 597
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-485-101-597

Query Match      5.0%; Score 218; DB 7; Length 1026;
Best Local Similarity 20.3%; Pred. No. 0.017;
Matches 178; Conservative 140; Mismatches 410; Indels 148; Gaps 33;

QY 32 GYILKNRQ-----IKLIAASVAAAFQAHAGLGLNIQSNLDPPFGSITVTGEEAKA 86
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 GYLRHHQEEVGELLGQIQSGSAAQQAQMAET-----RDLKCDVTSALREIRA 276

QY 87 LLGGSV---TVSEKGLTAKVHKGDKAVIASVSEQAVRDPVLFRICAGAVREYTAIL 143
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 QLEGHAVQSTLQSEBEWFRVLDRLSEAAKVNNTDAMRSQAQEEITEYRRLQARTTLEALK 336

QY 144 DPVGSVPTKLSALSDGKTHKTAFTAESOE--NQAKALRKT-----DKKDSANAA 192
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 STKSLERQSELED-----RHQDIASYQAIQQLDAELRNTKWMAAQLREYQDQLLN-- 390

QY 193 VKPAYNGKTHVTKGETVKQIAAAIRPKHLLTLEQVADALLKANPNVSAHGRLAGSVLHI 252
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 VKMALDIEIAAYRKLEGECECRIGFGPIPPSLPBG---LPKIPSVSTH----- 435

QY 253 PNLNRIKAEQPKQOTAKPKAETAMPSEPSQATVEKPEKPEAKVAAPAKAEKPAVRP 312
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 ---IKVKSSEKIKVVEKSEKETEVIVEEQTEETQVTEETVEEKEAKEEGKEEGGEE 492

QY 313 E-----PVPAAANTAASETAESAPOQAAASAIPTPTDETGNVAVSEPVQVSAEET 363
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 EAEGGEETKSPPAEAAASPEKEAKSPVKEEAKSPAEEKSPEKEAKS--PAE--VKSPKA 550

QY 364 ESSLFGSGVTLILAGGAALIALLLLRLAQSKKARRTEESVPE--EPPDLDDAADDDGIEI 422
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 KS-----PAKEAKSPPEAKSPEKEAKSP----- 575

QY 423 TFAVETPATPEPAPKNDVNDTIALDGESEEEISAKQTFDVTETPS---NRIDLDFSL 479
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 576 --AEVKSPEKAKSPAEEAKS---PAEAKSPEKAKSPVKEEAKSPAEEKSPEKEAKSP 629

QY 480 AAQNGLSALTQDE--ETOKRADADNNAIESTDSVYEPETFPN---YNPVIEVIDTPEP 535
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 630 AEVKSPEKAKSPVKEEAKSPAEEKSPEKAKSP---EAKSPEKEAKSPAEEKSPEKAKSPA 685

QY 536 ESWAQAENK--PETVDTDFSDNLPSNNHIGTBETASAKPASPSGLAGFLK--ASSPETIL 592
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 686 AKSPVKAASPEKAKSPAEEKSPEKAKSPAEEKAK--SPEKAKSPVKEEAKTPEKAK 743
```

593 EKTVAEQPEELHDFLKVVETDAVETAPETDFDFAAADLSALLQPAEAPSVERNITE 652
 744 SPVKEAKSPKAKSPKAKTLVDKSPKA-KTAKAEKSPADKFFPEKAKSPVKEVKSP 802
 653 TVAETDFDNATADDLSALLQSEVPVPAVEENAAEIVADDLSALLQPAEAP-AVEENVTTTV 711
 803 EKAKSP-----LKEDAKA--PEKEIPKKEEVKSPVKEEKPQEVKVPKPAEKEKAPAT 856
 712 AETSDFHATAADDLSALLQPAEVAENVTKTV-----AETPDNATADDLSALLQ-SE 765
 857 PKTEEKDSDSKE-EAPKKEAPKPKVEKEPAVEKPKESKVEAKKEAEADKKVPTPEKE 915
 766 VPA-----VEENA-----AETLETPDSNTSEADALPDFLKDGEETVDSVYLSE--ENIP 815
 916 APKVEKEDAKPKETEVAKKEPDDAKAEPKPAEKPAAEKKDTEKAKKPEKP 975
 816 NNADTSPSPSVGSDAPSEAKYDLAEMYLEIGDRDA 851
 976 KTEAKAKEDDKTLKSPKPKAEKAKESSTDDQKDS 1011

RESULT 8
 US-10-258-899A-1487
 ; Sequence 1487, Application US/10258899A
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Drmanac, Radoje T.
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Xu, Chongjun
 ; APPLICANT: Cao, Yicheng
 ; APPLICANT: Ma, Yunquing
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Wang, Zhi wei
 ; APPLICANT: Xue, Aidong
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Goodrich, Ryle
 ; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
 ; FILE REFERENCE: 787CIP2-2G/US
 ; CURRENT APPLICATION NUMBER: US/10/258,899A
 ; CURRENT FILING DATE: 2003-11-06
 ; PRIOR APPLICATION NUMBER: PCT/US01/04098
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 09/774,434
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: 09/728,422
 ; PRIOR FILING DATE: 2000-11-30
 ; PRIOR APPLICATION NUMBER: 09/693,325
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 09/663,561
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: 09/654,936
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 09/620,325
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/598,075
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: 09/560,875
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: 09/496,914
 ; PRIOR FILING DATE: 2000-02-03
 ; NUMBER OF SEQ ID NOS: 3960
 ; SOFTWARE: Custom
 ; SEQ ID NO 1487
 ; LENGTH: 1026
 ; TYPE: PRT

; ORGANISM: Homo sapiens
 US-10-258-899A-1487
 Query Match 4.9%; Score 215; DB 6; Length 1026;
 Best Local Similarity 20.7%; Pred. No. 0.023;
 Matches 181; Conservative 134; Mismatches 415; Indels 144; Gaps 34;
 32 GYLKXNRQ-----IKLIAVAVAASFOAHAGLGLINQSNLDEPFSGSITVTGEEAKA 86
 228 GYLRRHHQEEVGGELLQIQSGGAAQAQMAET-----RDALKCDVTSALREIRA 276
 87 LLOGGSV---TVSEKGLTAKVHKLGDKNVAVSSEQAVRDPVLVFRIGAGAOVREYTAIL 143
 277 QLESHAVQSTLQSEEEFVRDLRLSEAAKVNTDAMRSQAQEEITTEYRQLQARTTELEALK 336
 144 DPVGSPTKLSALSDGKTHKTKTAPTAEBSQE--NQNAKALRKT-----DKKDSANAA 192
 337 STKDSLERQSELED-----RHQADIASYQEIQQLODAELNWKWMAAQLREYQDLN-- 390
 193 VKPAYNGKTHVVRKGETVKQIAAAIRPKHJTLTQVADALLKAMPNVSAHGLRAGSVLHI 252
 391 VKMALDIEIAAYRKLLEGEERCICFGPIPFSLPEG-----LPKIPSVSTH----- 435
 253 PNLNRKAEQPKQTAKPKAETASMPSEPSKQATVEKPVKPEAKVAAPEAKAKPAVRP 312
 436 ---IKVSEEEKIKVVEKSEKETVIVEEQTEETQVTEETEEKEEKEEGEGEEE 492
 313 EPVPAANTAASATAESAPQEAASAIIDTPTDETGNVSE-----PVBQVSAE-EETESGL 367
 493 E-----AEGGEETKSPVVEEAS-----PEKEAKSPVKEEAKSPAEKSPKEAKSP- 541
 368 FGGSYTLLLAGGGAALLIALLLRLAQSKKARTEESVPVEEEDLDAAADDGITEITFAEV 427
 542 -----AEVKSPEKAKSPKEAKSPVKEEAK-----EA 572
 428 ETPA---TPPAPKNDVNDTLALDGESEBELSAQTDFDVTDTPS---NRIDLDFDSLAA 481
 573 KSPAEEVKSPEKA-KSPAEEAKSPAEAKSPKEAKSPVKEEAKSPAEKSPVKEEAKSPAE 631
 482 AONGILSGALTQDR-ETQKRADADWNAIESTDSVYEPETFPN---YNPVEIVDTPPEPS 537
 632 VKSPEKAKSPKKEAKSPKEAKSP-----ERAKSPKEEAKSPKEAKSPVKEEAKSPKAK 687
 538 VAQTAENK-PETVDTDFSDNLPSNNHIGTTEETAAPSPAGLAGFLK--ASSPETILEK 594
 688 SPVKAEEKSPKAKSPVKEEAKSPKAKSPVKEEAK--SPEKAKSPVKEEAKTPEKAKSP 745
 595 TVAEQVTPPEELHDFLKVVETDAVETAPETDFDFAAADLSALLQPAEAPSVVERNITETV 654
 746 VKEAKSPKAKSPKAKTLVDKSPKA-KTAKAEKSPADKFFPEKAKSPVKEEAKSPKAK 804
 655 AETPDNATADDLSALLQSEVPVPAVEENAAEIVADDLSALLQPAEAP-AVEENVTTTVAE 713
 805 AKSP-----LKEDAKA--PEKEIPKKEEVKSPVKEEKPQEVKVPKPAEKEKAPATPK 858
 714 TSDFHATAADDLSALLQPAEVAENVTKTV-----AETPDNATADDLSALLQ-SEVP 767
 859 TEEKDSDSKE-EAPKKEAPKPKVEKEPAVEKPKESKVEAKKEAEADKKVPTPEKAP 917
 768 A---VEENA-----AETLETPDSNTSEADALPDFLKDGEETVDSVYLSE--ENIPNN 817
 918 AKVEKEDAKPKETEVAKKEPDDAKAEPKPAEKPAAEKKDTEKAKKPEKPKT 977
 818 ADTSPSPSVGSDAPSEAKYDLAEMYLEIGDRDA 851
 978 EAKAKEDDKTLKSPKPKAEKAKESSTDDQKDS 1011

RESULT 9
 US-10-425-114A-72422
 ; Sequence 72422, Application US/10425114A
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114A
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 72422
LENGTH: 748
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 701128931_FIL.pap
US-10-425-114A-72422

Query Match 4.9%; Score 214; DB 6; Length 748;
Best Local Similarity 20.3%; Pred. No. 0.019;
Matches 156; Conservative 110; Mismatches 307; Indels 195; Gaps 31;
QY 150 PKTKSALSDGKTHRTAPTAEQONAKALRDKDSANAAYKPAYNGKTHTVRKGET 209
DB 22 PPSKSAKGGK--RQADEIEKQVSARKKTEEVAQKQKEAKVQKKKSSDDSSSESE 79
QY 210 VKQIAAIRPHHLTEQVADALLKANPNVSAHGLRAG--SVLHLPNLRKAEQPKPT 267
DB 80 EKPAKPVLSKKVPAKGA-ALPKKDKPVSSSDSDSDSDSDSDSDSDSDSDSDSD 138
QY 268 AKPAETASMPSEPSKQATVEKPEAKVAAPAEKAEKPAVPEPVAANTAASETAA 327
DB 139 TVQKNKXSSDDSSDSEDEKPV---AKAAVPSK-----TPAKSSNISTPAK 184
QY 328 ESAPQEAAGAIPTPTDETGNVSEPEVQSAEETBSGLFGGSYTLILLAGGGAALL 387
DB 185 KGKPASSSS--SDSSEDDSDSD----- 206
QY 388 LLLRLAQSKARRTEESVPEEPDLDDAADDGIBITFAEVETPATPPAP-----KNDV 441
DB 207 ---ELATKQNEVKVQKGESSDDSSD-----SEDEKPAKVAVPSKNGAKNGT 257
QY 442 NDTLALDGESEELSAKQTFDVTED---TPSNRIDLDFSLAAQNGILSGALQDEETQ 498
DB 258 LSTLAKKGKPAASSSSSDSDSDDEDEAPKTV---APAKNGHASTKKTOPSSES- 311
QY 499 KRADADNAIESTD---SVPEPTFN-PYNPVEIVIDTPEPSVAQTAENKPTVDTDF 553
DB 312 --SDSDSDSDSDSDDEGSKKKPTTAKLPLV-----APAKVSSDDES 355
QY 554 SDNLPNNHIGTEETASAKPASGLAGFLKASSPETILKTAEVQTPPELHDFLKVE 613
DB 356 SES-----SDNDKAPVATA-----VKPSARAKKVESDSDSDSDSD----- 394
QY 614 TDAVETAPETPFNAADLSDALLQPAZAPSVENITETVAETPDNATADDLSALLQ 673
DB 395 -----EDKMDIDDDSDSDE-----SEEPQKXAVKNSKESDSESEDESE-EKP 441
QY 674 SEVPVAVENAAEIVADDLSALLQPAE--APAVENVTETVAETSDPHTAADDLSALLQ 731
DB 442 SKTPQKRGDVMVMDAALSEKAPKPTVTPRESGTSKTL-----FVGNLPFSVERA 493
QY 732 EVPAVENVTKVAEIPDFNATADDLSALLQPEVPVAVENAA-----EIT 777
DB 494 DV-----EGFDFKAGEVVDVRFATDDTGKFGHGFVEFATAEAQAQNALGLNGQOLFNR 549
QY 778 LE-----TPDS-----NTSEADALPDFLKQGEETV---DWSIVLSEENIPNADTSF 822
DB 550 LDLARERGATPNSNWNNSQKS-----ERQSQITFVRGFDTSIGDEIRGSLQEHF 603
QY 823 PS-----ESVGDSPASEAKYDIAEMYLEIGDRDAAAETVQKLLERAE 864

DB 604 GSCGDITRVSIPIKDYESGAVKGFA--YVDFGDGDMGKALE--LHETE 647
US-10-687-046-5
RESULT 10
US-10-687-046-5
Sequence 5, Application US/10687046
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
FILE REFERENCE: A-59941-1/RFT/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/687,046
CURRENT FILING DATE: 2003-10-15
PRIOR APPLICATION NUMBER: US/10/080,505
PRIOR FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 1702
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-687-046-5
Query Match 4.8%; Score 213; DB 6; Length 1702;
Best Local Similarity 20.0%; Pred. No. 0.046;
Matches 165; Conservative 105; Mismatches 314; Indels 240; Gaps 31;
QY 38 NRQIKLTAASVAVAAASFOAHAGLGLNIQSLNDEPPSGSITVTGEEAKALIG---GGSVT 94
DB 744 NRNPK--ATNINVTNNATLYSGRNVANITSI-----TASDNKAKVHIGYKAGDTVC 792
QY 95 VSEK---GLTAKVHKLGDKAVIAVSSEQAVRDPVLV---FRIG-----AGAQV 136
DB 793 VRSDYTGVTCTTDLKSLDKALNSFNATVSGNVLSNANFVLKANLFGTISGTNSQV 852
QY 137 R-EYTAILDPVGYSPKTKSALSQDKTHRKTAPTAESQENQNAKALRK-----182
DB 853 RLTESSHHLTGDSDNVNQLNLDKXGHIHLNAQNDANKVTTNTLTVNSLSGNSFYLYTDL 912
QY 183 TDKKDSANAAYKAYNGKTHV--RKGETVKQTAARPKHLLTLEQVADALLKANPNVSA 240
DB 913 SNKQGDKVVTKSATGNFTLQVADKTGEPTK-----NELTDFDASNA-TRNNLVSL 963
QY 241 HG-----RLRAGSVLHLPNLRKAEQPKQTAKPKA 272
DB 964 VGNTVDLGAWKYKLRNVNGRYDLYNVEVEKRNQTVDTNTTTNNIQADVVP-----1014
QY 273 ETASMPSEPSKQATVEKPKVEKPEAKVAAPAEKAEKPAVPEPVAANTAASETAASAPQ 332
DB 1015 ---SVPSNNEEIARVETPV-----PPAPATPSETTETVAENSKQ 1051
QY 333 EAAA-----SAIDTDTDEGNVSEPEVQVSABEETESGLFGSYTLILLAGGGAALL 386
DB 1052 ESKTVKNEQDQATET-TAQNGEVAEEAKPSVKANTQTNVEAQSGET-----EETQTEI 1105
QY 387 LLLRLAQSKARR-RTEESVPEEPDLDDAADDGIBITFAEVETPATPPAPKNDVNDTL 445
DB 1106 KETAKVEKEKAKVEKEKAKVEKEIQEAPQ-----MASETSPKQAKPAKPEVSTDTK 1159
QY 446 ALDGE-----SEELSAKQTFDVTETDTPSNRIDLDFSLAAQNGILSGALTQ 493
DB 1160 VEETQVQAQPTQSTTTVAEEAATSPNSKPAEETQPSKTNABEVTVPVSKNQTEN---TT 1216
QY 494 DEETQKRAADNNAIBSTDSVTEPETFPNPNVPIVIDTPEPSVAQTAENKPTVDTDF 553
DB 1217 DQPTREKTA---KVETKTPBPQVQASQKQESQETVQQAQVLE-----1260
QY 554 SDNLPNNHIGTEETASAKPASGLAGFLKASSPETILKTAEVQTPPELHDFLKVE 613
DB 1261 SENVPTVNN-----AEEVQAQLOQ-----Q 1280

```

614  TDVAETAPETPPDFNAAADDLSALLOPAEAPSVEENITETVAETPD-----FNATADDLS 668
1281  TSATVSTKQAPENSINTGSATAITETAEKSDKPQ--TETAAS TEDASQHKANTVADNSV 1338
669  A-----LLQSEVPVPEENAA---EIVADDLSALLOPAEAPAVEENVTTET 710
1339  ANNESSEPKSRRRRSISQPSAEETTAATDETTIANSKRKNRERSRVRSEPT 1398
711  VAETSDPHTKAA-----DLSALLOPAEVP--VEENVTKTVAE 746
1399  VTNGSRSTVALRDLSTNTNAVISDAMAKAQFVALNVGRKVSQ 1442

RESULT 11
PCT-US03-27401-316
; Sequence 316, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 316
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-316

```

Query Match	4.8%; Score 213; DB 1; Length 1881;
Best Local Similarity	19.6%; Pred. No. 0.051;
Matches	164; Conservative 113; Mismatches 301; Indels 258; Gaps 37;
Qy	80 TGEAKALLGGSVTVS-----EKGLTAKVHKLGDKAVIAVSSQAVRDPVLVFRIGAG 133
Db	51 TGTEASVLTAFLGLTVGSLLIYKPKKTASVFLVGAMGLVLPISAGAV-DPVNTIALASR 109
Qy	134 AQVREYTA-----ILDVPGV-----SPKTSALSDGKTHR 163
Db	110 EGVMEGRRYGVYLSGDLTKLGLDVLLETSAKPGVTVVETPQSIITNQEQARTEN 169
Qy	164 KTAPTAESQENQAKALRKTDKOSANAAPVANGKTHTVRKGTVTVQIAAIPRKILT 223
Db	170 QWVETEEAPKEAPKT--EESPKEPKSEVKPT---DQTLTPKVBEGEDSAEPAP---- 219
Qy	224 LEQVADALLKANPNVSAHGRLRAGSVLHPIPNLRIKASQPKQTAKPAKAEATSMPESEPS 283
Db	220 VEEVG-----GEVESPEEKVAVPKESQSPDKPAEESK 252
Qy	284 QATVKEPKPEAKVAAPKAKEPAVRPEVPAA---NTAASET--AAESAPOFAAASA 338
Db	253 VEQAGEPV-----APR-EDEKAPVPEKQPEAPEEKAVEETPKQESTPTDKABET 303
Qy	339 IDTPTDETGN-AVSP-VEQVSAEETSGLFGGSYVTLILAGGAALIALLLLLLRAOSK 396
Db	304 VE-PKBEETVNSIQIEQPKVETPAVEKQTE----- 330
Qy	397 RARTEESVPSEEPDLDDAADDGIEITPAEVETPATPRAPKNDVNTLALDGESEELS 456
Db	331 -----PTBEPKVEQA-----GEPVAPREDEQAP-----TAPVEPEKQPEVP 366
Qy	457 AKQTFDVEDTDTESNRIDLDLFDLSAAQNGILSGALTQDBETOKRADADNALESTDSVYE 516
Db	367 EEKAVEETPKPEDKI-----KGIGTKEPKVSKSELNNQIDKA--SSVSPTD--YS 412
Qy	517 PETFPNPNV-----BIVDTTPBESVAQTAENKET-VDTDFSNLPNNHIGTEETAS 570
Db	413 TASYNALGVLETAQGVASPEPKQPEVNSEYNKILKTAIDALNVDKTELNNITADAKT- 471


```

; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1311
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7513018CD1
PCT-US03-24084-5

Query Match      4.8%; Score 210; DB 1; Length 1311;
Best Local Similarity 22.3%; Pred. No. 0.048;
Matches 161; Conservative 59; Mismatches 345; Indels 158; Gaps 30;

QY 153 KSALSDGKTHRKTAAPTAESENQNAKALRTDKKDSANAAVKAYNGKTHVKG----- 207
Db 201 KQSTGDKKETSAKETQSIKTSKOLAPTSK-----VLAKP--TPKAETTTKGPALTT 253
QY 208 -----ETVKQIAAAIRPKHITLQVADALLKANPNVSAHGLRAGSVL---HIPNLNRK 259
Db 254 PKEPTPTPKPEASTTKPTPTTKSA--FTTPKEPAPTTTKSAPTTPKPEAPTTTKPEP 311
QY 260 A-----EOPKQOTAKPKAETA--SMSPBPSKOA--TVEKPVKPKAKVAPEAKAEKPAVR 311
Db 312 APTTPKEPAPTTTKPEAPTTTKSAPTTKPEAPTTPKKPAPTTKPEAPTTKPEPTTTP 371
QY 312 PEPVPAANTAASETAEAFQEAASAIIDTDTGNAVSEPPVQVSAEBETESGLFGGS 371
Db 372 KEPAPTTKEPAPTTKPEAPTPAKKPAPTTKPEAPTTKPEAPTTTKPESTP----- 425
QY 372 YTLLAGGAALIALLLRLAQKRAARTEESVPEEPDLDAAADGGIITFAEYETPA 431
Db 426 -----PKEPAPTTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTKPEPSTT 463
QY 432 TPEPA---PKNDVNDTLALDGESEELSAKOTFOVETDTPSNRIDLDPSLAAQNGILS 488
Db 464 TKEPAPTTKPEAPTTPKKPAPTTKPEAPTTKPEAPTTKPEAPTTTK-----PAPTAPKE 514
QY 489 GALTQDEETQKRADADNNAISTDSVYE-----PETNPNVNPVIVIDTPE-----PE 536
Db 515 PAPTTPKET-----APTTPKKLITPTPEKLAFTTPEKPAFTTPELAPTTPEEPTTPTPE 569
QY 537 SVAQT-----AENKP-----ETVDTDSNLPSNNHIGTETASAKPSGLAGF 582
Db 570 EPAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTT-----PKETA--PTTPKGTGTA-- 620
QY 583 LKASSPETILE--KTVAEVQTPTELHDLFKVYETDAVAE-----TAPETDPFNAAA 631
Db 621 -----PTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKPAFTTPKGTAPTTKPEAPT 675
QY 632 DDL-----SALLQPA-----EAPSVEENTETVAETPDNATADLSALLQPS 675
Db 676 TPKEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKPAFTTPKGTAPTTKPEAPT 732
QY 676 -UPAVEENAAEIVADDLSALLQPAAPAVEENVETVA--ETSDFTHTAADDLSALLQPAEV 733
Db 733 TAPTTKPEPAPTTPKKPAFTTTPETPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAEP 792
QY 734 PAVEEN-----VTKT--VAEIPDNATADLSALLQPSVEPAVEENAAEITLTPDSN 784
Db 793 PKALENSKPEPGVPTTKTAAKPEMTTAKOTHTERDLRTTPTTAAAPKMTKET--AT 850
QY 785 TSE 787
Db 851 TTE 853

RESULT 15
PCT-US03-28227-4168
; Sequence 4168, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGEANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RHOX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4168
; LENGTH: 3897
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 958089.PT451p
PCT-US03-28227-4168

Query Match      4.8%; Score 209.5; DB 1; Length 3897;
Best Local Similarity 19.8%; Pred. No. 0.14;
Matches 161; Conservative 118; Mismatches 321; Indels 215; Gaps 32;

QY 145 PVGYSPTKSALSDGKTHRKTAAPTAESENQNAKALRTDKKDSANAAVKAYNGKTHV 204
Db 1811 PVSFSSKTEK-----HSPVSPSAKTERHSPASSSSKTEK---HSPVSPSTKTERHS- 1858
QY 205 RKGETVQIAAAIRPKHITLQVADALLKANPNVSAHGLRAGSVLHIPNLNRKAEOPK 264
Db 1859 -----FVSSTKTERHPPVSPSGK-----TDKRPVPS 1886
QY 265 PQTAK-----PKAETASMPSEPSKQATVEKPEV---KPEAKV-AAPEAKAEKPAVRPEFV 315
Db 1887 GRTEKHPVPSQRTKRLPVSPSGRTDKHPVSTAGTKTEKHLPVSPSGKTEK---QPPVS 1943
QY 316 PAANT-AASETAAESAPQEAASAIIDTDTGNAVSEPPVQVSAEB-----361
Db 1944 PTKSTERIEETMSVRELKMAFQSGQDPSKHKTLGLPEHKSQKQKQKQKVRKEKGP 2003
QY 362 -----ETESGLFGGSYTLLAGGAALIALLLRLAQSKRARRTE-----402
Db 2004 LTQREBAQKTENQTKRGQRLPVTG-----TABSRRGVRSVSSIGVKKEDAAG 2050
QY 403 -----ESVPEEP-----DLDDAADGGIITFAEYETPA 431
Db 2051 KEKVLSHKIPEVPQSVPEEESHRESEVPKEMADQGDMDLQISPRKSTSTDFSEV--- 2106
QY 432 TPEAPKNDVNDTLALDGESE-----ELSAKQTFDVTOTPPSNRIDLDPSLAAQNGILS 488
Db 2107 INQELEDNDKYYQCFRLSEETEKALHLDQVLTSPTNTTFFLDYMKDEFPLALSLQSGALD 2166
QY 489 GALTQDEETQKRADADNNAISTDSVYEPEPTNPNVNPVIVIDTPE--PESVA----- 539
```


Search completed: December 12, 2003, 17:37:27
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:33:12 ; Search time 47 Seconds
(without alignments)
2971.902 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 880

Sequence: 1 MPAGRLPRRCPMWTKFTDCT.....EEAGDVLKRAQAQELGI 880

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	880	100.0	880	21	AA1980.DAT.*
2	59	6.7	244	24	ABP78566
3	59	6.7	244	24	ABP80077
4	56	6.4	132	24	ABP79950
5	56	6.4	253	24	ABP78538
6	33	3.8	275	24	ABP78509
7	33	3.8	275	24	ABP79684
8	9	1.0	351	24	ABU07414
9	9	1.0	396	19	AAW64771
10	9	1.0	396	21	AA1980.DAT.*
11	9	1.0	396	22	AA1981.DAT.*
12	9	1.0	396	22	AA1982.DAT.*
13	8	0.9	396	22	AA1983.DAT.*
14	8	0.9	396	22	AA1984.DAT.*
15	8	0.9	396	22	AA1985.DAT.*
16	8	0.9	396	22	AA1986.DAT.*
17	8	0.9	396	22	AA1987.DAT.*
18	8	0.9	396	22	AA1988.DAT.*
19	8	0.9	396	22	AA1989.DAT.*
20	8	0.9	396	22	AA1990.DAT.*
21	8	0.9	396	22	AA1991.DAT.*
22	8	0.9	396	22	AA1992.DAT.*
23	8	0.9	396	22	AA1993.DAT.*
24	8	0.9	396	22	AA1994.DAT.*
25	8	0.9	396	22	AA1995.DAT.*
26	8	0.9	396	22	AA1996.DAT.*
27	8	0.9	396	22	AA1997.DAT.*
28	8	0.9	396	22	AA1998.DAT.*
29	8	0.9	396	22	AA1999.DAT.*
30	8	0.9	396	22	AA2000.DAT.*
31	8	0.9	396	22	AA2001.DAT.*
32	8	0.9	396	22	AA2002.DAT.*
33	8	0.9	396	22	AA2003.DAT.*
34	8	0.9	396	22	AA2004.DAT.*
35	8	0.9	396	22	AA2005.DAT.*
36	8	0.9	396	22	AA2006.DAT.*
37	8	0.9	396	22	AA2007.DAT.*
38	8	0.9	396	22	AA2008.DAT.*
39	8	0.9	396	22	AA2009.DAT.*
40	8	0.9	396	22	AA2010.DAT.*
41	8	0.9	396	22	AA2011.DAT.*
42	8	0.9	396	22	AA2012.DAT.*
43	8	0.9	396	22	AA2013.DAT.*
44	8	0.9	396	22	AA2014.DAT.*
45	8	0.9	396	22	AA2015.DAT.*

10	9	1.0	396	20	AA1980.DAT.*
11	9	1.0	396	22	AA1981.DAT.*
12	9	1.0	396	22	AA1982.DAT.*
13	8	0.9	396	22	AA1983.DAT.*
14	8	0.9	396	22	AA1984.DAT.*
15	8	0.9	396	22	AA1985.DAT.*
16	8	0.9	396	22	AA1986.DAT.*
17	8	0.9	396	22	AA1987.DAT.*
18	8	0.9	396	22	AA1988.DAT.*
19	8	0.9	396	22	AA1989.DAT.*
20	8	0.9	396	22	AA1990.DAT.*
21	8	0.9	396	22	AA1991.DAT.*
22	8	0.9	396	22	AA1992.DAT.*
23	8	0.9	396	22	AA1993.DAT.*
24	8	0.9	396	22	AA1994.DAT.*
25	8	0.9	396	22	AA1995.DAT.*
26	8	0.9	396	22	AA1996.DAT.*
27	8	0.9	396	22	AA1997.DAT.*
28	8	0.9	396	22	AA1998.DAT.*
29	8	0.9	396	22	AA1999.DAT.*
30	8	0.9	396	22	AA2000.DAT.*
31	8	0.9	396	22	AA2001.DAT.*
32	8	0.9	396	22	AA2002.DAT.*
33	8	0.9	396	22	AA2003.DAT.*
34	8	0.9	396	22	AA2004.DAT.*
35	8	0.9	396	22	AA2005.DAT.*
36	8	0.9	396	22	AA2006.DAT.*
37	8	0.9	396	22	AA2007.DAT.*
38	8	0.9	396	22	AA2008.DAT.*
39	8	0.9	396	22	AA2009.DAT.*
40	8	0.9	396	22	AA2010.DAT.*
41	8	0.9	396	22	AA2011.DAT.*
42	8	0.9	396	22	AA2012.DAT.*
43	8	0.9	396	22	AA2013.DAT.*
44	8	0.9	396	22	AA2014.DAT.*
45	8	0.9	396	22	AA2015.DAT.*

ALIGNMENTS

RESULT 1

AA1980.DAT.*
AA1981.DAT.*
ID AA1980.DAT.* standard; Protein: 880 AA.

XX AA1980.DAT.*

AC AA1980.DAT.*

DT 18-APR-2000 (first entry)

DE N. meningitidis T-cell stimulating protein A (TspA).

XX T-cell stimulating protein A; TspA; CD4+ T-cell; stimulant; meningitis;
XX antibacterial; anti-inflammatory; vaccine; neisserial disease;
XX gonorrhea; septicemia; septic arthritis; pelvic inflammatory disease;
XX meningococcal; gonococcal.

OS Neisseria meningitidis.

XX WO200003003-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-GB02205.

PR 10-JUL-1998; 98GB-0014902.

PA (UYN0-) UNIV NOTTINGHAM.

PI Ala'Aldeen D, Todd I;

DR WPI; 2000-147612/13.

DR N-PSDB; AA249702.

XX

PT Generation of cell lines and clones specific to a particular protein
PT for screening antigenic peptides which are used as vaccines in treating
XX meningococcal, gonococcal infections
XX
XX
XX Claim 59; Page 42-45; 51pp; English.
XX
XX The present sequence is N. meningitidis (strain SD, serogroup B
CC (B:15:P1.16)) T-cell stimulating protein A (Tspa). Tspa is a CD4+ T-cell
CC stimulant. It can be produced recombinantly using lambda ZapII phage
CC library comprising the DNA encoding Tspa. Tspa has antibacterial and
CC anti-inflammatory activity and can be used in vaccine formulations
CC against neisserial diseases like meningitis, gonorrhoea, septicaemia,
CC septic arthritis and pelvic inflammatory diseases. T-cell lines and
CC clones specific to neisserial proteins can be generated for screening
CC meningococcal or gonococcal genomic phage display libraries to
CC identify peptides which stimulate T-cell lines and clones.
XX
XX Sequence 880 AA;
SQ
Query Match 100.0%; Score 880; DB 21; Length 880;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGRLPRRCPMWTKFTDCTRSNRIQPPTHRGYILKNNRQIKLIAASVAAAFQAHAGL 60
DB 1 MPAGRLPRRCPMWTKFTDCTRSNRIQPPTHRGYILKNNRQIKLIAASVAAAFQAHAGL 60
QY 61 GGLNIQSNLDPPFGSGITVTGEERKALIGGSGVTVSEKGLTAKVHKGDKAVIASSQQA 120
DB 61 GGLNIQSNLDPPFGSGITVTGEERKALIGGSGVTVSEKGLTAKVHKGDKAVIASSQQA 120
QY 121 VRDPVLVFRIGAGQVREYTAILOPVGVSPKTSALSGDKTKRKTAPTAEQENQAKAL 180
DB 121 VRDPVLVFRIGAGQVREYTAILOPVGVSPKTSALSGDKTKRKTAPTAEQENQAKAL 180
QY 181 RKTDKDSANAARVPAAYNGKTHTVRKGTQVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
DB 181 RKTDKDSANAARVPAAYNGKTHTVRKGTQVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRLRAGSVLHHPNLNRIKASQPKQTAQKAEATSMFSEPSKQATVEKPEKPAKVA 300
DB 241 HGRLRAGSVLHHPNLNRIKAEQPKQTAQKAEATSMFSEPSKQATVEKPEKPAKVA 300
QY 301 PEAKAEKPAVRPEPVPAAATASETAASAPQEAASALDPTDGTGNVSPVQVSAE 360
DB 301 PEAKAEKPAVRPEPVPAAATASETAASAPQEAASALDPTDGTGNVSPVQVSAE 360
QY 361 BETESGLFGGSYTLILAGGGAALIALLLRLAQSKRARRTESVPEEPDLDAAADGGI 420
DB 361 BETESGLFGGSYTLILAGGGAALIALLLRLAQSKRARRTESVPEEPDLDAAADGGI 420
QY 421 EITFAEVTPTATPEAPKNDVNDTIALDGSESEELSAKQTFVETDTPSNRIDLDPSLA 480
DB 421 EITFAEVTPTATPEAPKNDVNDTIALDGSESEELSAKQTFVETDTPSNRIDLDPSLA 480
QY 481 AAQNGILSALGTDSEETQKRAADWNAIESTSVTEPETFNPVPEIVIDTPEESVAQ 540
DB 481 AAQNGILSALGTDSEETQKRAADWNAIESTSVTEPETFNPVPEIVIDTPEESVAQ 540
QY 541 TAENKPEVTDTDFSDNLPSNNHIGTEETASAKPASPSGLAGFKASSPTILEKTVAEVQ 600
DB 541 TAENKPEVTDTDFSDNLPSNNHIGTEETASAKPASPSGLAGFKASSPTILEKTVAEVQ 600
QY 601 TPEELHDFLKVYETDAVETAPETPDFNAAADLSALLOPAEAPSVEENITETVAETPDF 660
DB 601 TPEELHDFLKVYETDAVETAPETPDFNAAADLSALLOPAEAPSVEENITETVAETPDF 660
QY 661 NATADLSALLOPSEVPVPAVEENNAEIVADDLSALLOPAEAPAVEENVETVAETSFHTA 720
DB 661 NATADLSALLOPSEVPVPAVEENNAEIVADDLSALLOPAEAPAVEENVETVAETSFHTA 720
QY 721 ADDLSALLOPAEVPVPAVEENVTKTVAEIPDFNATADLSALLOPSEVPVPAVEENNAEITLET 780
DB 721 ADDLSALLOPAEVPVPAVEENVTKTVAEIPDFNATADLSALLOPSEVPVPAVEENNAEITLET 780

DB 721 ADDLSALLOPAEVPVPAVEENVTKTVAEIPDFNATADLSALLOPSEVPVPAVEENNAEITLET 780
QY 781 PDNNTSEADALPDFLKDGEETVDWSIYLSEENIPNNADTFPSESYGSDAPSAKYDLA 840
DB 781 PDNNTSEADALPDFLKDGEETVDWSIYLSEENIPNNADTFPSESYGSDAPSAKYDLA 840
QY 841 EMYLEIGDRDAAAEETVOKLLEAEAGDVLKRAQALAEGLI 880
DB 841 EMYLEIGDRDAAAEETVOKLLEAEAGDVLKRAQALAEGLI 880
RESULT 2
ABP78566
ID ABP78566 standard; Protein; 244 AA.
XX
AC ABP78566;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 3662.
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
FN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX
DR WPI; 2003-058415/05.
XX
PT N-PSDB; ABZ39536.
XX
PS New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX
PS Disclosure; Page 453; 815pp; English.
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 244 AA;
Query Match 6.7%; Score 59; DB 24; Length 244;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 822 FPSESVGSDAPSEAKYDLAEYMLEIGDRDAAAEETVOKLLEAEAGDVLKRAQALAEGLI 880
DB 186 FPSESVGSDAPSEAKYDLAEYMLEIGDRDAAAEETVOKLLEAEAGDVLKRAQALAEGLI 244
RESULT 3
ABP80077
ID ABP80077 standard; Protein; 244 AA.
XX
AC ABP80077;
XX
DT 07-MAR-2003 (first entry)
XX

DE N. gonorrhoeae amino acid sequence SEQ ID 6684.
XX Antibacterial; infection; vaccine; gene therapy.
KW Neisseria gonorrhoeae.
OS WO200279243-A2.
XX 10-OCT-2002.
PD 12-FEB-2002; 2002WO-IB02069.
PF 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
PA Fontana MR, Pizza M, Masignani V, Monaci E;
PI WPI; 2003-058415/05.
DR N-PSDB; ABZ41047.
DR New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 674; 815pp; English.
PS The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 244 AA;
Query Match 6.7%; Score 59; DB 24; Length 244;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 822 FPSSVSGSDAPSEAKYDLAEMYLEIGDRDAAATVQKLEBAEGDVLKRAQALAEGLI 880
DB 186 FPSSVSGSDAPSEAKYDLAEMYLEIGDRDAAATVQKLEBAEGDVLKRAQALAEGLI 244
RESULT 4
ABP79950
ID ABP79950 standard; Protein; 132 AA.
AC ABP79950;
XX 07-MAR-2003 (first entry)
DT N. gonorrhoeae amino acid sequence SEQ ID 6430.
DE Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
OS WO200279243-A2.
XX 10-OCT-2002.
PD 12-FEB-2002; 2002WO-IB02069.
PF 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
PA Fontana MR, Pizza M, Masignani V, Monaci E;
PI WPI; 2003-058415/05.
DR

DR N-PSDB; ABZ40920.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 656; 815pp; English.
PS The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 132 AA;
Query Match 6.4%; Score 56; DB 24; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.9e-44;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 449 GESEELSAKQTFDVTDPNSRIDLPDSLAAQNGILSGALTQDEETOKRADAD 504
DB 77 GESEELSAKQTFDVTDPNSRIDLPDSLAAQNGILSGALTQDEETOKRADAD 132
RESULT 5
ABP78538
ID ABP78538 standard; Protein; 253 AA.
AC ABP78538;
XX 07-MAR-2003 (first entry)
DT N. gonorrhoeae amino acid sequence SEQ ID 3606.
DE Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
OS WO200279243-A2.
XX 10-OCT-2002.
PD 12-FEB-2002; 2002WO-IB02069.
PF 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
PA Fontana MR, Pizza M, Masignani V, Monaci E;
PI WPI; 2003-058415/05.
DR N-PSDB; ABZ39508.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 450; 815pp; English.
PS The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 253 AA;
Query Match 6.4%; Score 56; DB 24; Length 253;

[illegible]

XX Novel genes which are differentially regulated in prostate cancer,
PT useful for diagnosing prostate cancer in prostate tissue sample and
PT assessing therapeutic or preventive intervention in prostate cancer
PT patients -
XX
XX
XX Claim 1; Page 244-245; 416pp; English.
XX
XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity.
CC (I) is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type
CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (I) can be used as target for therapy or drug
CC discovery. (I) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (I) is
CC useful in therapeutic applications to treat prostate cancer. The
CC identification of specific genes, and groups of genes, expressed in
CC pathways physiologically relevant to prostate cancer permits the
CC definition of functional and disease pathways and the delineation of
CC targets in these pathways which are useful in diagnostic, therapeutic,
CC and clinical applications. This is the amino acid sequence of a protein
CC differentially regulated in prostate cancer.
XX
XX SQ Sequence 351 AA;
XX
XX Query Match 1.0%; Score 9; DB 24; Length 351;
XX Best Local Similarity 100.0%; Pred. No. 16;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 385 ALLLLRLA 393
XX | | | | | | | |
XX Db 175 ALLLLRLA 183
XX
XX RESULT 9
XX AAW64771
XX ID AAW64771 standard; Protein; 396 AA.
XX AC
XX AC AAW64771;
XX DT 11-NOV-1998 (first entry)
XX DE
XX DE Serine threonine kinase VRK1.
XX KW Serine threonine kinase; VRK1; VRK2; BIR kinase; cell growth control;
XX KW antitumour agent.
XX OS Homo sapiens.
XX XX WO9829552-A1.
XX PN
XX XX 09-JUL-1998.
XX PD
XX XX

PF 25-DEC-1997; 97WO-JP04855.
XX
PR 27-DEC-1996; 96JP-0357864.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX Nezu J, Oku A;
XX
XX WPI; 1998-388133/33.
XX N-PSDB; AAV46338.
XX Serine-threonine kinase highly expressed in actively growing cells -
PT useful for development of cell growth inhibitors and antitumour
PT agents
XX
XX Claim 1; Page 21-24; 63pp; Japanese.
XX
XX This sequence is the serine threonine kinase VRK1 of the invention.
XX The protein is highly expressed in actively growing cells (such
XX as foetal organs), and has significant homology with the BIR kinase of
XX vaccinia virus. The kinase can be used as a substrate for the
XX screening of potential inhibitors, which can then be used in the control
XX of cell growth and as antitumour agents. Antisense DNA delivered via a
XX suitable vector can also be used for control of cell growth.
XX
XX SQ Sequence 396 AA;
XX
XX Query Match 1.0%; Score 9; DB 19; Length 396;
XX Best Local Similarity 100.0%; Pred. No. 18;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 824 SESVGSADP 832
XX | | | | | | | |
XX Db 59 SESVGSADP 67
XX
XX RESULT 10
XX AAY13484
XX ID AAY13484 standard; Protein; 396 AA.
XX AC
XX AC AAY13484;
XX DT 26-JUL-1999 (first entry)
XX DE
XX DE Human VRK1 kinase (NVRK1).
XX KW VRK1 kinase; NVRK1; cell proliferation disorder; immune response;
XX KW cancer; AIDS; Addison's disease; allergy; atherosclerosis; lupus;
XX KW atopic dermatitis; diabetes mellitus; multiple sclerosis; infection;
XX KW rheumatoid arthritis; osteoarthritis; osteoporosis; pancreatitis;
XX KW autoimmune thyroiditis; adenocarcinoma; leukaemia; lymphoma; melanoma;
XX KW sarcoma; teratocarcinoma.
XX OS Homo sapiens.
XX XX
XX PN WO9911794-A2.
XX XX
XX PD 11-MAR-1999.
XX XX
XX PF 04-SEP-1998; 98WO-US18524.
XX XX
XX PR 04-SEP-1997; 97US-0923469.
XX XX
XX (INCY-) INCYTE PHARM INC.
XX Corley NC, Lal P, Yue H;
XX WPI; 1999-337367/28.
XX N-PSDB; AAX55605.
XX
XX New VRK1 kinase useful for treating disorders associated with cell
XX proliferation and cancer

PS Claim 1; Fig 1A-E; 63pp; English.

XX This represents a new human VRK1 kinase (NVRK1). Host cells containing
CC a vector comprising the nucleic acid are used for the recombinant
CC production of the protein. NVRK1 is useful to diagnose, prevent, and
CC treat, disorders associated with cell proliferation. The VRK1 antagonist
CC is used to treat cancer or an immune response. Immune response may be
CC associated with disorders such as AIDS, Addison's disease, allergies,
CC atherosclerosis, atopic dermatitis, diabetes mellitus, multiple
CC sclerosis, lupus, rheumatoid arthritis, osteoarthritis, osteoporosis,
CC pancreatitis, autoimmune thyroiditis, complications of cancer,
CC haemodialysis, and extracorporeal circulation, viral, bacterial, fungal,
CC parasitic, protozoal, and helminthic infections, and trauma. Cancers
CC include adenocarcinoma, leukaemia, lymphoma, melanoma, sarcoma, and
CC teratocarcinoma.

XX SQ Sequence 396 AA;

Query Match 1.0%; Score 9; DB 20; Length 396;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 SESVGSADP 832
Db 59 SESVGSADP 67

RESULT 11

AAG67433
ID AAG67433 standard; Protein: 396 AA.

XX AC AAG67433;

XX DT 26-NOV-2001 (first entry)

XX DE Amino acid sequence of a human polypeptide.

XX KW Human; protein kinase; protein phosphatase; signal transduction;
XX intracellular signalling pathway.

XX OS Homo sapiens.

XX PN WO200109345-A1.

XX PD 08-FEB-2001.

XX PF 28-JUL-2000; 2000WO-JP05060.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 18-OCT-1999; 99US-0159590.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 17-FEB-2000; 2000US-0183322.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
PI Senoo C, Nezu J;

XX DR WPI; 2001-564736/63.

XX New genes encoding protein kinase and protein phosphatase, useful for
PT identifying modulators which can be used to treat human or animal
PT disorders associated with the expression or function of these enzymes -

XX Example 4; Page 190-193; 336pp; Japanese.

XX The specification describes human protein kinase/protein phosphatases.
CC The polypeptides are expected to participate in signal transduction
CC in cells. The kinase phosphatases are connected with intracellular
CC signalling pathways. Antisense oligonucleotides and compounds
CC identified by screening (agonists or antagonists) can be used to

CC treat human or animal disorders associated with the expression
CC or function of the protein. In addition, the polypeptides may be used
CC as target molecules for drug development. The present sequence
CC represents a polypeptide, used in the course of the invention.

XX SQ Sequence 396 AA;

Query Match 1.0%; Score 9; DB 22; Length 396;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 SESVGSADP 832
Db 59 SESVGSADP 67

RESULT 12

AAG67612
ID AAG67612 standard; Protein: 396 AA.

XX AC AAG67612;

XX DT 26-NOV-2001 (first entry)

XX DE Amino acid sequence of a human protein.

XX KW Human; protein kinase; protein phosphatase; signal transduction.

XX OS Homo sapiens.

XX PN WO200109316-A1.

XX PD 08-FEB-2001.

XX PF 28-JUL-2000; 2000WO-JP05061.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 18-OCT-1999; 99US-0159590.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 17-FEB-2000; 2000US-0183322.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
PI Senoo C, Nezu J;

XX DR WPI; 2001-570286/64.

XX New genes encoding proteins with protein kinase/protein phosphatase
PT activity, useful in the diagnosis and treatment of diseases -

XX Example 4; Page 88-91; 233pp; Japanese.

XX The specification describes human protein kinase/protein phosphatases.
CC It is expected that the protein kinase/protein phosphatase gene
CC participates in signal transduction in cells. The protein
CC kinase/protein phosphatase polypeptides and polynucleotides are
CC useful for developing diagnostics and treatment agents for human
CC and animal diseases. The protein kinase/protein phosphatase polypeptides
CC are useful as target molecules in designing novel drugs. The protein
CC kinase/protein phosphatase polynucleotides are useful as a source of
CC probes and primers, which may be used to isolate homologous sequences.
CC The present sequence represents a human protein, which is used in the
CC course of the invention.

XX SQ Sequence 396 AA;

Query Match 1.0%; Score 9; DB 22; Length 396;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 SESVGSQAP 832
Db |||||
59 SESVGSQAP 67

RESULT 13
AAU29895
ID AAU29895 standard; Protein; 57 AA.
XX
AC AAU29895;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #386.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
XX Claim 20; Page 207; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation, to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemia. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 57 AA;

Query Match 0.9%; Score 8; DB 22; Length 57;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 KHLTLEQV 227
Db |||||
27 KHLTLEQV 34

RESULT 14
AAG57437

ID XX
AC XX
AC AAG57437;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 74021.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.9%; Score 8; DB 21; Length 70;
Best Local Similarity 100.0%; Pred.No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 LIAALLLL 390
DB 7 LIAALLLL 14

RESULT 15
AAG61518
ID AAG61518 standard; Protein; 70 AA.
XX
AC AAG61518;
XX
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 79800.
DE

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139452.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 22-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

Mon Dec 15 09:21:33 2003

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.9%; Score 8; DB 21; Length 70;
Best Local Similarity 100.0%; Pred. No: 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 LIALLLLL 390
Db 7 LIALLLLL 14

Search completed: December 12, 2003, 17:39:12
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:38:23 ; Search time 22 Seconds
(without alignments)
1692.434 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 880

Sequence: 1 MPAGELPRCPMTWKFTDCT.....EEAEGDVLKRAQALQELGI 880

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	9	1.0	107	3	US-09-102-528-23
2	9	1.0	107	3	US-09-102-528-27
3	9	1.0	396	2	US-08-878-989-16
4	9	1.0	396	3	US-08-272-796-16
5	9	1.0	396	3	US-09-344-700-2
6	8	0.9	122	4	US-09-252-991A-21614
7	8	0.9	157	4	US-09-252-991A-19220
8	8	0.9	363	4	US-09-252-991A-31330
9	8	0.9	402	4	US-09-252-991A-19220
10	8	0.9	524	4	US-09-252-991A-17388
11	8	0.9	929	4	US-09-252-991A-22946
12	8	0.9	21	3	US-08-848-580-6
13	7	0.8	21	5	PCT-US92-04537-2
14	7	0.8	30	1	US-08-117-083-25
15	7	0.8	37	3	US-09-177-249-118
16	7	0.8	40	4	US-09-480-993-15
17	7	0.8	51	3	US-08-818-112-104
18	7	0.8	51	4	US-08-818-111-99
19	7	0.8	51	4	US-09-056-556-104
20	7	0.8	51	4	US-09-072-596-99
21	7	0.8	95	2	US-08-465-640-2
22	7	0.8	142	1	US-07-956-700B-54
23	7	0.8	142	1	US-08-476-537-54
24	7	0.8	142	1	US-08-485-607-54
25	7	0.8	142	2	US-08-475-879-54
26	7	0.8	142	4	US-09-433-043B-54
27	7	0.8	143	1	US-07-956-700B-43

28 7 0.8 143 1 US-08-476-537-43 Sequence 43, Appl
29 7 0.8 143 1 US-08-485-607-43 Sequence 43, Appl
30 7 0.8 143 2 US-08-475-879-43 Sequence 43, Appl
31 7 0.8 143 4 US-09-433-043B-43 Sequence 43, Appl
32 7 0.8 145 4 US-09-134-001C-5194 Sequence 5194, Ap
33 7 0.8 150 4 US-09-252-991A-29265 Sequence 29265, A
34 7 0.8 152 4 US-09-187-999-2 Sequence 2, Appl
35 7 0.8 165 1 US-07-956-700B-106 Sequence 106, App
36 7 0.8 165 1 US-08-476-537-106 Sequence 106, App
37 7 0.8 165 1 US-08-485-607-106 Sequence 106, App
38 7 0.8 165 2 US-08-475-879-106 Sequence 106, App
39 7 0.8 165 4 US-09-433-043B-106 Sequence 106, App
40 7 0.8 170 4 US-09-433-043B-127 Sequence 127, App
41 7 0.8 171 4 US-09-252-991A-217874 Sequence 217874, A
42 7 0.8 180 4 US-09-252-991A-21755 Sequence 21755, A
43 7 0.8 181 4 US-09-252-991A-16898 Sequence 16898, A
44 7 0.8 182 4 US-09-325-932A-80 Sequence 80, Appl
45 7 0.8 183 4 US-09-252-991A-26371 Sequence 26371, A

ALIGNMENTS

RESULT 1
US-09-102-528-23
; Sequence 23, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..107
; OTHER INFORMATION: /note= "Figure 9, sequence of
; Zm41a"
US-09-102-528-23

Query Match 1.0%; Score 9; DB 3; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 VRKGETVKQ 212
Db 8 VRKGETVKQ 16
|||||

RESULT 2
US-09-102-528-27
; Sequence 27, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA: US/09/102,528
; FILING DATE:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-102-528-27

Query Match 1.0%; Score 9; DB 3; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 VRKGETVKQ 212
Db 8 VRKGETVKQ 16

RESULT 3
US-08-878-989-16
; Sequence 16, Application US/08879989
; Patent No. 5885803
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,989
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1827450
; US-09-102-528-27

Query Match 1.0%; Score 9; DB 3; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1827450
; US-08-878-989-16

Query Match 1.0%; Score 9; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 SESVGSADP 832
Db 59 SESVGSADP 67

RESULT 4
US-09-272-796-16
; Sequence 16, Application US/09272796
; Patent No. 6207148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/272,796
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/878,989
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1827450
; US-09-272-796-16

Query Match 1.0%; Score 9; DB 3; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 824 SESVGSAP 832

Db 59 SESVGSAP 67
|||||

RESULT 5

US-09-344-700-2

; Sequence 2, Application US/09344700

; Patent No. 6265194

; GENERAL INFORMATION:

; APPLICANT: Nezu, Jun-Ichi

; APPLICANT: Oku, Asuka

; TITLE OF INVENTION: NOVEL SERINE-THREONINE KINASE GENE

; FILE REFERENCE: 06501/033001

; CURRENT APPLICATION NUMBER: US/09/344,700

; CURRENT FILING DATE: 1999-06-25

; EARLIER APPLICATION NUMBER: PCT/JP97/0485

; EARLIER FILING DATE: 1997-12-25

; EARLIER APPLICATION NUMBER: JP 8/357864

; EARLIER FILING DATE: 1996-12-27

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 396

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-344-700-2

Query Match

Best Local Similarity 1.0%; Score 9; DB 3; Length 396;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 SESVGSAP 832

Db 59 SESVGSAP 67
|||||

RESULT 6

US-09-252-991A-21614

; Sequence 21614, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21614

; LENGTH: 122

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21614

Query Match

Best Local Similarity 0.9%; Score 8; DB 4; Length 122;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 476 FDSLAAQ 483

Db 88 FDSLAAQ 95
|||||

RESULT 7

US-09-252-991A-19220

; Sequence 19220, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 19220

; LENGTH: 157

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-19220

Query Match

Best Local Similarity 0.9%; Score 8; DB 4; Length 157;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAGRLPRR 9

Db 104 PAGRLPRR 111
|||||

RESULT 8

US-09-252-991A-31330

; Sequence 31330, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31330

; LENGTH: 363

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31330

Query Match

Best Local Similarity 0.9%; Score 8; DB 4; Length 363;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 EQVADALL 232

Db 310 EQVADALL 317
|||||

RESULT 9

US-09-292-097-16

; Sequence 16, Application US/09292097B

; Patent No. 632977

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Kaser, Matthew, R.

; APPLICANT: Baughn, Mariah, R.

; TITLE OF INVENTION: TAPASIN-LIKE PROTEIN

; FILE REFERENCE: PC-0002 US

; CURRENT APPLICATION NUMBER: US/09/292,097B

; CURRENT FILING DATE: 1999-04-14

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PERL Program

; SEQ ID NO 16

;; LENGTH: 402
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE: -
;; OTHER INFORMATION: 3183699
US-09-292-097-16

Query Match 0.9%; Score 8; DB 4; Length 402;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 LLLAGGGA 381
Db 7 LLLAGGGA 14

RESULT 10
US-09-252-991A-17388
; Sequence 17388, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17388
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17388

Query Match 0.9%; Score 8; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAGRLPRR 9
Db 29 PAGRLPRR 36

RESULT 11
US-09-252-991A-22946
; Sequence 22946, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22946
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (691)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-22946

Query Match 0.9%; Score 8; DB 4; Length 929;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 QPAEAPAV 703
Db 533 QPAEAPAV 540

RESULT 12
US-08-848-580-6
; Sequence 6, Application US/08848580
; Patent No. 6013619
; GENERAL INFORMATION:
; APPLICANT: Cochran, Charles G
; APPLICANT: Revak, Susan D
; TITLE OF INVENTION: NOVEL PULMONARY SURFACTANTS AND
; TITLE OF INVENTION: THERAPEUTIC USES, INCLUDING PULMONARY LAVAGE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6013619th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/848,580
; FILING DATE: 28-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/488,123
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/419,824
; FILING DATE: 11-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/060,833
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,397
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/293,201
; FILING DATE: 04-JAN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/141,200
; FILING DATE: 06-JAN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 147.5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-784-2937
; TELEFAX: 619-784-9399
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-848-580-6

Query Match 0.8%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 LLLLLL 392

Db 5 LLLLLL 11

RESULT 13
PCT-US92-04537-2
; Sequence 2, Application PC/TU9204537
; GENERAL INFORMATION:
; APPLICANT: Cochran, Charles G
; APPLICANT: Revak, Susan D
; TITLE OF INVENTION: SYNTHETIC PULMONARY SURFACTANT PEPTIDES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 North Torrey Pines Road, Mail Drop TPC8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04537
; FILING DATE: 19920601
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,397
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bingham, Douglas A
; REGISTRATION NUMBER: 32,457
; REFERENCE/DOCKET NUMBER: SCR1025P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US92-04537-2

Query Match 0.8%; Score 7; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 386 LLLLLL 392
Db 5 LLLLLL 11

RESULT 14
US-08-117-083-25
; Sequence 25, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Boursnell, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA

Zip: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-58783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-117-083-25

Query Match 0.8%; Score 7; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 DMSIYLS 810
Db 10 DMSIYLS 16

RESULT 15
US-09-177-249-118
; Sequence 118, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; TITLE OF INVENTION: Development in Plants
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 118
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-177-249-118

Query Match 0.8%; Score 7; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 SQENQNA 177
Db 4 SQENQNA 10

Mon Dec 15 09:21:43 2003

us-09-743-674-2.olig.ra1

Page 6

Search completed: December 12, 2003, 17:41:54
Job time : 22 secs

1

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:40:49 ; Search time 36 Seconds

(without alignments)
4546.267 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 880

Sequence: 1 MPAGSLPRRCPMTKFTDCT.....EEAEGDLKRAQAQLGLGI 880

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 684280 seqs, 185983659 residues

Word size : 0

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
9: /cgn2_6/ptodata/2/pubpaa/US05A_PUBCOMB.pcp:*
10: /cgn2_6/ptodata/2/pubpaa/US09E_PUBCOMB.pcp:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	9	1.0	396	12	US-09-769-970-16
2	9	1.0	396	12	US-10-434-588-2
3	9	1.0	396	15	US-10-059-585-31
4	8	0.9	193	12	US-10-238-075-861
5	8	0.9	243	12	US-10-291-253A-14
6	8	0.9	399	12	US-10-293-971-9
7	8	0.9	402	10	US-09-933-561-16
8	8	0.9	503	15	US-10-103-140-2
9	8	0.9	532	9	US-09-833-790-428
10	8	0.9	734	14	US-10-008-355-5
11	8	0.9	1332	15	US-10-239-316-9
12	7	0.8	9	10	US-09-916-201-13
13	7	0.8	12	15	US-10-101-499-6
14	7	0.8	13	10	US-09-813-333-59
15	7	0.8	13	14	US-10-044-703-59

Sequence 80, Appl
Sequence 60, Appl
Sequence 61, Appl
Sequence 60, Appl
Sequence 61, Appl
Sequence 6, Appl
Sequence 40, Appl
Sequence 194, App
Sequence 245, App
Sequence 247, App
Sequence 118, App
Sequence 118, App
Sequence 15, Appl
Sequence 1984, Ap
Sequence 104, App
Sequence 99, Appl
Sequence 33, Appl
Sequence 3368, A
Sequence 5469, Ap
Sequence 1, Appl
Sequence 195, Appl
Sequence 14264, A
Sequence 7592, Ap
Sequence 19, Appl
Sequence 627, App
Sequence 5347, Ap
Sequence 33689, A
Sequence 6522, Ap
Sequence 80, Appl
Sequence 14, Appl

15 12 US-10-079-167-80
20 10 US-09-813-333-60
20 10 US-09-813-333-61
20 14 US-10-044-703-60
20 7 0.8
21 8 US-10-044-703-61
21 8 US-08-488-123-6
21 14 US-10-038-612-40
22 7 0.8
23 9 US-09-739-907-194
24 7 0.8
25 11 US-09-892-877-245
26 7 0.8
27 10 US-09-948-783-247
28 7 0.8
29 15 US-10-213-512-118
30 7 0.8
31 15 US-10-044-967-15
32 50 10 US-09-764-877-1984
33 51 12 US-10-084-843-104
34 51 12 US-10-193-002-99
35 51 12 US-10-098-732A-33
36 72 9 US-09-864-761-33368
37 93 15 US-10-106-698-5469
38 95 10 US-09-805-427A-1
39 106 9 US-09-739-907-195
40 118 15 US-10-156-761-14264
41 126 15 US-10-156-761-7592
42 128 15 US-10-311-111-19
43 131 9 US-09-925-297-627
44 136 15 US-10-106-698-5347
45 164 12 US-10-029-386-33689
167 10 US-09-738-626-6522
182 15 US-10-219-220-80
186 12 US-10-453-478-14

ALIGNMENTS

RESULT 1

US-09-769-970-16
; Sequence 16, Application US/09769970
; Publication No. US20030170219A1
; GENERAL INFORMATION:
; APPLICANT: Handman, Olga
; Corley, Neil C.
; Guegler, Karl G.
; Lal, Preeti
; Goli, Surya K.
; Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; KINASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/769,970
; FILING DATE: 24-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/272,796
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US

Mon Dec 15 09:21:47 2003

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 396 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 1827450
 SEQUENCE DESCRIPTION: SEQ ID NO: 16:
 US-09-769-970-16

Query Match 1.0%; Score 9; DB 12; Length 396;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 824 SESVGSAP 832
 Db 59 SESVGSAP 67

RESULT 2
 US-10-434-588-2
 ; Sequence 2, Application US/10434588
 ; Publication No. US20030171557A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nezu, Jun-Ichi
 ; TITLE OF INVENTION: NOVEL SERINE-THREONINE KINASE GENE
 ; FILE REFERENCE: 06501-033002
 ; CURRENT APPLICATION NUMBER: US/10/434,588
 ; CURRENT FILING DATE: 2003-05-09
 ; PRIOR APPLICATION NUMBER: US/09/563,997
 ; PRIOR FILING DATE: 2000-05-03
 ; PRIOR APPLICATION NUMBER: PCT/JP97/04855
 ; PRIOR FILING DATE: 1997-12-25
 ; PRIOR APPLICATION NUMBER: JP 8-357864
 ; PRIOR FILING DATE: 1996-12-27
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 396
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-434-588-2

Query Match 1.0%; Score 9; DB 12; Length 396;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 824 SESVGSAP 832
 Db 59 SESVGSAP 67

RESULT 3
 US-10-059-585-31
 ; Sequence 31, Application US/10059585
 ; Publication No. US20030082776A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ota, Toshio
 ; APPLICANT: Isogai, Takao
 ; APPLICANT: Nishikawa, Tetsuo
 ; APPLICANT: Hayashi, Koji
 ; APPLICANT: Otsuka, Kaoru
 ; APPLICANT: Yamamoto, Jun-ichi
 ; APPLICANT: Ishii, Shizuko
 ; APPLICANT: Sugiyama, Tomoyasu
 ; APPLICANT: Wakamatsu, Ai

APPLICANT: Nagai, Keiichi
 APPLICANT: Otsuki, Tetsuji
 APPLICANT: Funahashi, Shin-Ichi
 APPLICANT: Senoo, Chiaki
 APPLICANT: Nezu, Jun-Ichi
 TITLE OF INVENTION: NOVEL GENES ENCODING PROTEIN
 TITLE OF INVENTION: KINASE/PROTEIN PHOSPHATASE
 FILE REFERENCE: 06501-098001
 CURRENT APPLICATION NUMBER: US/10/059,585
 CURRENT FILING DATE: 2002-01-29
 PRIOR APPLICATION NUMBER: PCT/JP00/05060
 PRIOR FILING DATE: 2000-07-28
 PRIOR APPLICATION NUMBER: US 60/183,322
 PRIOR FILING DATE: 2000-02-17
 PRIOR APPLICATION NUMBER: US 60/159,590
 PRIOR FILING DATE: 1999-10-18
 PRIOR APPLICATION NUMBER: JP 2000-118776
 PRIOR FILING DATE: 2000-01-11
 PRIOR APPLICATION NUMBER: JP 2000-183767
 PRIOR FILING DATE: 2000-05-02
 PRIOR APPLICATION NUMBER: JP 11-248036
 PRIOR FILING DATE: 1999-07-29
 NUMBER OF SEQ ID NOS: 64
 SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 31
 ; LENGTH: 396
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-059-585-31

Query Match 1.0%; Score 9; DB 15; Length 396;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 824 SESVGSAP 832
 Db 59 SESVGSAP 67

RESULT 4
 US-10-238-075-861
 ; Sequence 861, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated from the genome of the bacterium Escherichia coli and biological uses of these polynucleotides and of the bacterium Escherichia coli
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 861
 ; LENGTH: 193
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-10-238-075-861

Query Match 0.9%; Score 8; DB 12; Length 193;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 ALLLRL 392
 Db 42 ALLLRL 49

RESULT 5
 US-10-291-253A-14
 ; Sequence 14, Application US/10291253A
 ; Publication No. US20030150017A1

```

; GENERAL INFORMATION:
; APPLICANT: Botella, Jose
; APPLICANT: Graham, Michael
; APPLICANT: Fairbairn, David
; TITLE OF INVENTION: A Method for Facilitating Pathogen Resistance
; FILE REFERENCE: nematode
; CURRENT APPLICATION NUMBER: US/10/291,253A
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: PR8706
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: PR8802
; PRIOR FILING DATE: 2001-11-12
; PRIOR APPLICATION NUMBER: US60/341404
; PRIOR FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Meloidogyne incognita
US-10-291-253A-14

```

```

Query Match      0.9%; Score 8; DB 12; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      374 LLLAGGGA 381
      |||||
Db     120 LLLAGGGA 127

```

```

RESULT 6
US-10-293-971-9
; Sequence 9, Application US/10293971
; Publication No. US20030172408A1
; GENERAL INFORMATION:
; APPLICANT: CHEN, RUOYING
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: DA COSTA E SILVA, OSWALDO
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS OF USE IN PLANTS
; FILE REFERENCE: 16313-0177
; CURRENT APPLICATION NUMBER: US/10/293,971
; CURRENT FILING DATE: 2001-11-12
; PRIOR APPLICATION NUMBER: 60/344,503
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-10-293-971-9

```

```

Query Match      0.9%; Score 8; DB 12; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      323 SETAESA 330
      |||||
Db     381 SETAESA 388

```

```

RESULT 7
US-09-933-561-16
; Sequence 16, Application US/09933561
; Patent No. US20020106664A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti G.
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: TAPASIN-LIKE PROTEIN
; FILE REFERENCE: PC-0002-1 CIP

```

```

; CURRENT APPLICATION NUMBER: US/09/933,561
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 09/292,097
; PRIOR FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 16
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 3183699
US-09-933-561-16

```

```

Query Match      0.9%; Score 8; DB 10; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      374 LLLAGGGA 381
      |||||
Db     7 LLLAGGGA 14

```

```

RESULT 8
US-10-103-140-2
; Sequence 2, Application US/10103140
; Publication No. US20030113817A1
; GENERAL INFORMATION:
; APPLICANT: LI, Li
; APPLICANT: LI, V. Yang
; TITLE OF INVENTION: Hemogen-EDAG: No. US20030113817A1el Nuclear Factors Expressed in I
; TITLE OF INVENTION: Development
; FILE REFERENCE: 38368-179726
; CURRENT APPLICATION NUMBER: US/10/103,140
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 60/277,624
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-103-140-2

```

```

Query Match      0.9%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      357 VSAEEETE 364
      |||||
Db     102 VSAEEETE 109

```

```

RESULT 9
US-09-833-790-428
; Sequence 428, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodon
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428

```

```
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-790-428

Query Match      0.9%; Score 8; DB 9; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      478 SLAAQNG 485
Db      40 SLAAQNG 47
|||||

RESULT 10
US-10-008-355-5
; Sequence 5, Application US/10008355
; Publication No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.0040101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5

Query Match      0.9%; Score 8; DB 14; Length 734;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      47 SVAAASF 54
Db      23 SVAAASF 30
|||||

RESULT 11
US-10-239-316-9
; Sequence 9, Application US/10239316
; Publication No. US20030125253A1
; GENERAL INFORMATION:
; APPLICANT: TANIYAMA, Yoshio
; APPLICANT: KITA, Shunbun
; APPLICANT: SATOMI, Tomoko Komiyama
; TITLE OF INVENTION: No. US20030125253A1el Protein, Process for Producing The Same And
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/10/239,316
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 9
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Human
US-10-239-316-9

Query Match      0.9%; Score 8; DB 15; Length 1332;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      388 LLLRLAQS 395
|||||

us-09-743-674-2.olig.rapb

Db      14 LLLRLAQS 21
|||||

RESULT 12
US-09-916-201-13
; Sequence 13, Application US/09916201
; Patent No. US20020131976A1
; GENERAL INFORMATION:
; APPLICANT: LALVANI, Ajit
; APPLICANT: PATHAN, Ansar A.
; APPLICANT: HILL, Adrian V.S.
; TITLE OF INVENTION: TUBERCULOSIS VACCINE
; FILE REFERENCE: 117-359
; CURRENT APPLICATION NUMBER: US/09/916,201
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 09/467,893
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/113,783
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 9
; TYPE: PRT
; ORGANISM: M.tuberculosis
US-09-916-201-13

Query Match      0.8%; Score 7; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      333 EAAASAI 339
Db      3 EAAASAI 9
|||||

RESULT 13
US-10-101-499-6
; Sequence 6, Application US/10101499
; Publication No. US20030049654A1
; GENERAL INFORMATION:
; APPLICANT: Dahiyat, Bassil I.
; APPLICANT: Hayes, Robert J.
; APPLICANT: Bentzien, Joerg
; APPLICANT: Fiebig, Klaus M.
; TITLE OF INVENTION: PROTEIN DESIGN AUTOMATION FOR PROTEIN LIBRARIES
; FILE REFERENCE: A-67229-6/RPT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/10/101,499
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: US/09/782,004
; PRIOR FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US 60/181,630
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: US 60/186,904
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: US 60/197,851
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-101-499-6

Query Match      0.8%; Score 7; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 44;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      384 IALLLLL 390
|||||
```

```
Db      |||||
        4 IALLLL 10

RESULT 14
US-09-813-333-59
; Sequence 59, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-59

Query Match      0.8%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 333 EAAASAI 339
Db      |||||
        6 EAAASAI 12

RESULT 15
US-10-044-703-59
; Sequence 59, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-59

Query Match      0.8%; Score 7; DB 14; Length 13;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 333 EAAASAI 339
Db      |||||
        6 EAAASAI 12

Search completed: December 12, 2003, 17:46:30
Job time : 37 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:37:33 ; Search time 27 Seconds

(without alignments)
3134.387 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 880

Sequence: 1 MPAGRLPRRCPCMTKFTDCT.....EEAEGDVLKRAQALAEUGI 880

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168692 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	880	100.0	880	2	G81786
2	374	42.5	875	2	C81209
3	9	1.0	288	2	A87650
4	9	1.0	351	2	G01430
5	9	1.0	452	2	A69237
6	9	1.0	756	2	S74742
7	8	0.9	87	2	T26867
8	8	0.9	99	2	E83244
9	8	0.9	131	2	B72460
10	8	0.9	158	2	T49567
11	8	0.9	193	2	B91170
12	8	0.9	193	2	B86016
13	8	0.9	200	2	T09783
14	8	0.9	242	2	G82206
15	8	0.9	247	2	A12581
16	8	0.9	247	2	G97363
17	8	0.9	261	2	S30437
18	8	0.9	308	2	E72627
19	8	0.9	360	2	G83321
20	8	0.9	368	2	E69796
21	8	0.9	430	1	D70985
22	8	0.9	458	2	A75386
23	8	0.9	499	1	A29412
24	8	0.9	516	2	T04590
25	8	0.9	521	1	HYBSN
26	8	0.9	521	2	JQ2129
27	8	0.9	543	2	G87598
28	8	0.9	652	2	B82724
29	8	0.9	655	2	C84560

30 8 0.9 749 2 A75560 conserved hypothet
31 8 0.9 791 2 H72552 hypothetical prote
32 8 0.9 916 2 G82993 probable ATP-bind
33 8 0.9 1238 2 T03465 probable exonuclea
34 8 0.9 1321 2 T00382 hypothetical prote
35 7 0.8 65 2 A42131 hypothetical prote
36 7 0.8 69 2 S72828 hypothetical prote
37 7 0.8 75 2 A34445 H+-transporting tw
38 7 0.8 77 2 H96955 hypothetical prote
39 7 0.8 79 2 E86803 prophage p13 prote
40 7 0.8 81 2 T03705 reverse transcript
41 7 0.8 90 2 B83542 hypothetical prote
42 7 0.8 95 2 A70803 early secretory an
43 7 0.8 107 2 A2489 hypothetical prote
44 7 0.8 108 2 A70689 hypothetical prote
45 7 0.8 110 2 A13496 methyltransferase

ALIGNMENTS

RESULT 1

G81786

Neisseria-specific antigen protein, TspA NMA2146 [imported] - Neisseria meningitidis (str)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G81786

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel;
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: G81786

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-880 <PAR>

A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAH85358.1; PID:g738076;

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: tspA; NMA2146

Query Match	100.0%	Score	880;	DB	2;	Length	880;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	880;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MPAGRLPRRCPCMTKFTDCTRSNR	IQPPTHRGYLKNRQIKLIAASVAVAA	SFOAHAGL	60		
Db	1	MPAGRLPRRCPCMTKFTDCTRSNR	IQPPTHRGYLKNRQIKLIAASVAVAA	SFOAHAGL	60		
Qy	61	GGLNIQSNLDEPFGSGITVTGEEAK	ALLGGGVTVSEKGLTAKVHKLGDKAVI	AVSSEA	120		
Db	61	GGLNIQSNLDEPFGSGITVTGEEAK	ALLGGGVTVSEKGLTAKVHKLGDKAVI	AVSSEA	120		
Qy	121	VRDPVLVFRIGAGQVREYTAILDV	GVSPKTSALSDGKTHRTAPTASQEN	QNAKAL	180		
Db	121	VRDPVLVFRIGAGQVREYTAILDV	GVSPKTSALSDGKTHRTAPTASQEN	QNAKAL	180		
Qy	181	RKTDKDSANAAVKPAYNGKHTVR	KGETVKQIAAAIRPKHLTLEQVAD	ALLKANPNVSA	240		
Db	181	RKTDKDSANAAVKPAYNGKHTVR	KGETVKQIAAAIRPKHLTLEQVAD	ALLKANPNVSA	240		
Qy	241	HGRLRAGSVLHINLNRIKAEQPK	QPKOTAKPKABTASMPSEPSKOAT	VEKPEKPEAKVAA	300		
Db	241	HGRLRAGSVLHINLNRIKAEQPK	QPKOTAKPKABTASMPSEPSKOAT	VEKPEKPEAKVAA	300		
Qy	301	PEAKSKPAVRPEPVPAAANTA	ASETAESAPOAAASAIPTPTDET	GNVSEVEQVSAE	360		
Db	301	PEAKSKPAVRPEPVPAAANTA	ASETAESAPOAAASAIPTPTDET	GNVSEVEQVSAE	360		
Qy	361	EETESGLFGGSYTLILLAGGGA	LIALLLLRLAQSKRARTTESVPEE	EPDLDDAADDGI	420		
Db	361	EETESGLFGGSYTLILLAGGGA	LIALLLLRLAQSKRARTTESVPEE	EPDLDDAADDGI	420		

QY 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFDVETDTPSNRIDLDFDLSA 480
Db 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFDVETDTPSNRIDLDFDLSA 480
QY 481 AAQNGILSGALTODEETQKRADADWNAIESTDSVYEPETFPNPNVEIVIDTPEPESVAQ 540
Db 481 AAQNGILSGALTODEETQKRADADWNAIESTDSVYEPETFPNPNVEIVIDTPEPESVAQ 540
QY 541 TAENKPEVTDVDFSDNLPNNHHGTBETASAKPASPGLAGFLKASPETILEKTVAEVQ 600
Db 541 TAENKPEVTDVDFSDNLPNNHHGTBETASAKPASPGLAGFLKASPETILEKTVAEVQ 600
QY 601 TPEIHDFLKVYETDAVETAPEDPNAADDLALLQPAEAPSVENITETVAETPDF 660
Db 601 TPEIHDFLKVYETDAVETAPEDPNAADDLALLQPAEAPSVENITETVAETPDF 660
QY 661 NATADDLALLQSEVPAPVAVENAAEIVADDLALLQPAEAPSVENITETVAETSDPHTA 720
Db 661 NATADDLALLQSEVPAPVAVENAAEIVADDLALLQPAEAPSVENITETVAETSDPHTA 720
QY 721 ADDLSALLQPAEAPVAVENAAEIVADDLALLQPAEAPSVENITETVAETSDPHTA 780
Db 721 ADDLSALLQPAEAPVAVENAAEIVADDLALLQPAEAPSVENITETVAETSDPHTA 780
QY 781 PDNNTSEADALPDLKDGEBETVDMWSIYLSEENIPNNADTSFPPSESVGSDAPSEAKYDLA 840
Db 781 PDNNTSEADALPDLKDGEBETVDMWSIYLSEENIPNNADTSFPPSESVGSDAPSEAKYDLA 840
QY 841 EMYLEIGDRDAARTVQKLEAEAGDVLKRAQALAEELGI 880
Db 841 EMYLEIGDRDAARTVQKLEAEAGDVLKRAQALAEELGI 880

RESULT 2
C81209
tspA protein NMB0341 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81209
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ric, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, J.C.; Nelson, K.E.; Eisen, J.A.; Heidelberg, J.; Haft, D.H.; Kohn, M.; Salzberg, S.L.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81209
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-875 <TET>
A:Cross-references: GB:AB002391; GB:AB002098; NID:g7225561; PIDN:AAF40784.1; PID:g7225561
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0341

Query Match 42.5%; Score 374; DB 2; Length 875;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 674; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MPAGRLPRCPMTWKFTDCTRSNRIOPTPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
Db 1 MPAGRLPRCPMTWKFTDCTRSNRIOPTPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
QY 61 GGLNQTQSLNDEPFGSIVTVEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSFQA 120
Db 61 GGLNQTQSLNDEPFGSIVTVEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIASSFQA 120
QY 121 VRDPLVFRIGAGAVREYVAILDPVGYSPKTSALSGKTHRKTAAPTAAESQENQAKAL 180
Db 121 VRDPLVFRIGAGAVREYVAILDPVGYSPKTSALSGKTHRKTAAPTAAESQENQAKAL 180
QY 181 RKTDKKDSANAAPKPAYNGKTHTVRKGETVKQIAAARPKHLTLEQVADALLKANPNVSA 240

Db 181 RKTDKKDSANAAPKPAYNGKTHTVRKGETVKQIAAARPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRLRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKEPEAKVAA 300
Db 241 HGRLRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKEPEAKVAA 300
QY 301 PEAKAEKPAVRPEPVPAAANTAASETAESAPOAAAAAIDTPTDTCGNVSEPEVQVSAE 360
Db 301 PEAKAEKPAVRPEPVPAAANTAASETAESAPOAAAAAIDTPTDTCGNVSEPEVQVSAE 360
QY 361 EETES-----GLFGSGYTLLLAGGGAIALLLLLRLAQSKEARTEESVPEEPDLDAA 416
Db 361 EETESGLFDGLFGSGYTLLLAGGGAIALLLLLRLAQSKEARTEESVPEEPDLDAA 420
QY 417 DGGIETFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFDVETDTPSNRIDLDF 476
Db 417 DGGIETFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFDVETDTPSNRIDLDF 480
QY 477 DSLAAQNGILSGALTODEETQKRADADWNAIESTDSVYEPETFPNPNVEIVIDTPEPE 536
Db 477 DSLAAQNGILSGALTODEETQKRADADWNAIESTDSVYEPETFPNPNVEIVIDTPEPE 540
QY 537 SVAQTAENKPEVTDVDFSDNLPNNHHGTBETASAKPASPGLAGFLKASPETILEKTV 596
Db 537 SVAQTAENKPEVTDVDFSDNLPNNHHGTBETASAKPASPGLAGFLKASPETILEKTV 600
QY 597 ABOVTPPELHDFLKVYETDAVETAPEDPNAADDLALLQPAEAPSVENITETVAE 656
Db 597 ABOVTPPELHDFLKVYETDAVETAPEDPNAADDLALLQPAEAPSVENITETVAE 660
QY 657 TPDFNATADDLALLQPS 674
Db 657 TPDFNATADDLALLQPS 678

RESULT 3
A87650
hypothetical protein CC3235 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87650
R:Nierman, W.C.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kohn, M.; Salzberg, S.L.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87650
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <STO>
A:Cross-references: GB:AB005673; NID:g13424919; PIDN:AAK25197.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3235

Query Match 1.0%; Score 9; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 AALIALLL 389
Db 128 AALIALLL 136

RESULT 4
G01430
PL6 protein - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G01430
R:Duh, F.
submitted to the EMBL Data Library, May 1994
A:Reference number: H00612

A:Accession: G01430
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-351 <DUH>
 A:Cross-references: EMBL:U09584; NID:gl209019; PIDN:AAA92281.1; PID:gl209020
 C:Genetics:
 A:Gene: PL6

Query Match 1.0%; Score 9; DB 2; Length 351;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 ALLLLRLA 393
 |||||
 Db 175 ALLLLRLA 183

RESULT 5

A69297
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Jun-1999
 C:Accession: A69297
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: A69297
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-452 <KLE>
 A:Cross-references: GB:AE001078; GB:AE000782; NID:G2689401; PIDN:AAB90859.1; PID:G265025
 C:Superfamily: corrinoid/iron-sulfur protein small chain

Query Match 1.0%; Score 9; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 730 PAEYPAVEE 738
 |||||
 Db 95 PAEYPAVEE 103

RESULT 6

S74742
 exopolysaccharide export protein - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein sl10923
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S74742
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S74742
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-756 <KAN>
 A:Cross-references: EMBL:D90901; GB:AB001339; NID:gl651897; PIDN:BAAL6893.1; PID:d101762
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: epsB

Query Match 1.0%; Score 9; DB 2; Length 756;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 375 LIAGGGAAL 383
 |||||
 Db 464 LIAGGGAAL 472

RESULT 7

T26867
 hypothetical protein Y43F8C.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26867
 R:Ainscough, R.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20279
 A:Accession: T26867
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-87 <WIL>
 A:Cross-references: EMBL:AL032637; PIDN:CAA21608.1; CESP:Y43F8C.1
 A:Experimental source: clone Y43F8C
 C:Genetics:
 A:Gene: CESP:Y43F8C.1
 A:Introns: 51/1

Query Match 0.9%; Score 8; DB 2; Length 87;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 LIALLLLL 390
 |||||
 Db 4 LIALLLLL 11

RESULT 8

E83244
 conserved hypothetical protein PA3202 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83244
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brj
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: E83244
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <STO>
 A:Cross-references: GB:AE004744; GB:AE004091; NID:G9949317; PIDN:AAG06590.1; GSPDB:GN001;
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3202

Query Match 0.9%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 FDSLAAAQ 483
 |||||
 Db 65 FDSLAAAQ 72

RESULT 9

B72450
 Probable vacuolar ATP synthase subunit APE2326 - Aeropyrum pernix (strain Kl)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: B72450
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jinno, K.; Takaha
 awa H.; Takamiya, M.; Masuda, S.; Funanashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ku
 DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: B72460
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-131 <KAW>
 A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81338.1; PID:g1045124; PID:g510
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE2326

Query Match 0.9%; Score 8; DB 2; Length 131;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 LIALLLLL 390
 |||||
 DB 122 LIALLLLL 129

RESULT 10
 T49567
 related to attachment protein [imported] - Neurospora crassa
 N;Alternate names: protein B208.150
 C;Species: Neurospora crassa
 C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C;Accession: T49567
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A;Reference number: Z25022
 A;Accession: T49567
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-158 <SCH>
 A;Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.150
 A;Experimental source: BAC clone B208; strain OR74A
 C;Genetics:
 A;Gene: NCSP:B208.150
 A;Map position: 6

Query Match 0.9%; Score 8; DB 2; Length 158;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 PKPQTAKP 270
 |||||
 DB 41 PKPQTAKP 48

RESULT 11
 B91170
 hypothetical membrane protein [imported] - Escherichia coli (strain O157:H7, substrain R
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: B91170
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: B91170
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-193 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA837753.1; PID:g13363804; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs4330

Query Match 0.9%; Score 8; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 ALLLLRL 392
 |||||
 DB 42 ALLLLRL 49

RESULT 12

B86016
 hypothetical protein Z4855 [imported] - Escherichia coli (strain O157:H7, substrain EDL
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: B86016
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamousis, K.; Apodaca
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: B86016
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-193 <STO>
 A;Cross-references: GB:AE005174; NID:g12518143; PIDN:AAG58590.1; GSPDB:GN00145; UWGP:Z4
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z4855

Query Match 0.9%; Score 8; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 ALLLLRL 392
 |||||
 DB 42 ALLLLRL 49

RESULT 13

T09783
 dehydration-inducible homeobox leucine zipper protein Hb-1 - Craterostigma plantagineum
 C;Species: Craterostigma plantagineum
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
 C;Accession: T09783
 R;Frank, W.; Phillips, J.; Salamini, F.; Bartels, D.
 Plant J. 15, 413-421, 1999
 A;Title: Two dehydration-inducible transcripts from the resurrection plant Craterostigm
 A;Reference number: Z16852
 A;Accession: T09783
 A;Status: preliminary; translated from GB/EMBL/DBDJ
 A;Molecule type: mRNA
 A;Residues: 1-200 <FRA>
 A;Cross-references: EMBL:AJ005820; NID:e1294847; PID:e1294849
 C;Genetics:
 A;Gene: hb-1
 C;Keywords: DNA binding; stress-induced protein; transcription factor; transcription re

Query Match 0.9%; Score 8; DB 2; Length 200;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 PAEAPAVE 704
 |||||
 DB 2 PAEAPAVE 9

RESULT 14

G82206
 hypothetical protein VCL1375 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C;Accession: G82206
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
 1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: G82206

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-242 <HEI>

A;Cross-references: GB:AE004217; GB:AE003852; NID:g9555866; PIDN:AAF94533.1; GSPDB:GN001

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC1375

A;Map position: 1

Query Match 0.9%; Score 8; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 TVEKPVVK 293

Db 83 TVEKPVVK 90

RESULT 15

AI2581

hypothetical protein hisA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C;Accession: AI2581

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: A82577; MUID:21608550; PMID:11743193

A;Accession: AI2581

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-247 <KUR>

A;Cross-references: GB:AE008688; PIDN:AAL41071.1; PID:gl7738359; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: hisA

A;Map position: circular chromosome

C;Superfamily: N-(5'-phospho-D-riboseylformimino)-5-amino-1-(5''-phosphoribosyl)-4-imidaz

Query Match 0.9%; Score 8; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AVRDPVLV 127

Db 106 AVRDPVLV 113

Search completed: December 12, 2003, 17:41:20
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 17:33:43 ; Search time 18 Seconds
(without alignments)
2299.083 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 880

Sequence: 1 MPAGRLPRPCMTKFTDCT.....EEAGDVLKRAQALAEIGI 880

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9	1.0	350	1	PL6_MOUSE	QWUHL mus musculus
2	9	1.0	351	1	PL6_HUMAN	Q12893 homo sapien
3	8	0.9	247	1	HIS4_AGR15	P58790 agrobacteri
4	8	0.9	261	1	HME3_BRARE	P31533 brachydanio
5	8	0.9	430	1	C139_MYCTU	O86330 mycobacteri
6	8	0.9	500	1	CBP3_WHAT	P15151 triticum ae
7	8	0.9	521	1	NPRE_BACAM	P06832 bacillus am
8	8	0.9	521	1	NPRE_BACSU	P06142 bacillus su
9	8	0.9	530	1	ZIC2_MOUSE	Q62520 mus musculus
10	8	0.9	532	1	ZIC2_HUMAN	Q95409 homo sapien
11	8	0.9	1238	1	SBCC_RHOCA	O68032 rhodobacter
12	7	0.8	60	1	HMEB_MXGL	P31536 myxine glut
13	7	0.8	94	1	ESA6_MYCTU	Q57165 mycobacteri
14	7	0.8	113	1	RLAL_DICDI	P22684 dictyosteli
15	7	0.8	118	1	YJ91_ARCFU	O28288 archaeoglob
16	7	0.8	124	1	YLA2_LACAC	P29471 lactobacill
17	7	0.8	156	1	BCP3_MOUSE	Q35290 mus musculus
18	7	0.8	168	1	IPYR_METAC	Q8tm3 methanosarc
19	7	0.8	169	1	IPYP_METMA	Q8pwy5 methanosarc
20	7	0.8	182	1	RLS5_THETH	P24315 thermus the
21	7	0.8	183	1	TLPI_CHICK	Q9YGV8 gallus gall
22	7	0.8	186	1	TLPI_HUMAN	O95753 homo sapien
23	7	0.8	186	1	TLPI_XENLA	Q9w622 xenopus lae
24	7	0.8	196	1	LAST_SERMA	P37006 serratia ma
25	7	0.8	201	1	HAM1_RALSO	Q8xxf4 ralstonia s
26	7	0.8	201	1	Y2F9_VIBCH	Q9kxp3 vibrio chol
27	7	0.8	203	1	HEPT_MYCAV	P96794 mycobacteri
28	7	0.8	206	1	YOG1_YERPE	Q8zax0 yersinia pe
29	7	0.8	232	1	YHCG_BACSU	P54591 bacillus su
30	7	0.8	236	1	28KD_MYCLE	P19361 mycobacteri
31	7	0.8	242	1	SGAE_MYCPN	P75289 mycoplasma
32	7	0.8	243	1	CYL_EUGGR	P20114 euglena gra
33	7	0.8	258	1	ELI_HUMAN	Q9un11 homo sapien

34	7	0.8	260	1	IMB_XENLA	P52297 xenopus lae
35	7	0.8	261	1	ZNUB_ECOLI	P39832 escherichia
36	7	0.8	262	1	CUT3_MYCTU	O06318 mycobacteri
37	7	0.8	262	1	YA23_METJA	Q58429 methanococc
38	7	0.8	264	1	COMT_RAT	P22734 rattus norv
39	7	0.8	266	1	ELI_FIG	P00772 sus scrofa
40	7	0.8	266	1	ELI_RAT	P00772 rattus norv
41	7	0.8	267	1	YATR_BACFI	P26946 bacillus fi
42	7	0.8	280	1	TONB_NEIMA	P57003 neisseria m
43	7	0.8	280	1	TONB_NEIMB	P57004 neisseria m
44	7	0.8	295	1	SNA4_BOVIN	P81125 bos taurus
45	7	0.8	295	1	SNA4_HUMAN	P54920 homo sapien

ALIGNMENTS

RESULT 1
ID PL6_MOUSE STANDARD; PRT; 350 AA.
AC Q9WUHL; Q91V16;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE PL6 protein (Placental protein 6).
GN PL6 OR PP6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Ivanov S., Minna J., Lerman M.I.;
RT "Mouse ortholog of the human gene PL6."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
FN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Brapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Butterfield Y.S.N., Grimwood J., Schmutz J., Myers R.M.,
Rodríguez A.C., Krzywinski M.I., Skalska U., Smallus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF134238; AAD24193.1; -
CC DR EMBL; BC009099; AAH09099.1; ALT_INT.
CC DR EMBL; BC019473; AAH19473.1; -

DR MGD; MGI:1930765; P16.
 KW Transmembrane. 12 34 POTENTIAL.
 FT TRANSMEM 38 60 POTENTIAL.
 FT TRANSMEM 72 94 POTENTIAL.
 FT TRANSMEM 98 120 POTENTIAL.
 FT TRANSMEM 127 149 POTENTIAL.
 FT TRANSMEM 164 183 POTENTIAL.
 FT TRANSMEM 190 208 POTENTIAL.
 FT TRANSMEM 223 245 POTENTIAL.
 SQ SEQUENCE 350 AA; 9DB9B164CA2842D CRC64;
 Query Match 1.0%; Score 9; DB 1; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 385 ALLLLRLA 393
 DB 175 ALLLLRLA 183
 RESULT 2
 PL6 HUMAN
 ID PL6 HUMAN STANDARD; PRT; 351 AA.
 AC Q12893; Q14568; Q9UIX3.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE P16 protein (placental protein 6).
 GN P16 OR LUCA1.2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Latif F., Duh F.-M., Bader S., Lerman M.I., Minna J.D.;
 RT "A novel human cDNA that is homozygously deleted in small cell lung
 cancer and located to 3p21.3";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Barlow K., Miller N., Kramer J., Elliott G., Keppler D.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Muscle;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=20535986; PubMed=11085536;
 RA Lerman M.I., Minna J.D.;

"The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes";
 Cancer Res. 60:6116-6133 (2000).
 -|- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -|- TISSUE SPECIFICITY: Expressed strongly in kidney and skeletal muscle, followed by liver, placenta, pancreas, and lung, with low amounts in heart and only traces in brain.
 CC -|- CAUTION: Ref.2 (AAB67308) sequence differs from that shown due to erroneous gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to a license@isb-sib.ch).
 CC
 CC
 DR EMBL; U09584; AAA92281.1; -;
 DR EMBL; Z84492; CAB41768.1; -;
 DR EMBL; AC002481; AAB67308.1; ALT_SEQ.
 DR EMBL; BC011948; AAH11948.1; -;
 DR EMBL; BC017367; AAH17367.1; -;
 DR PIR; G01430; G01430.
 DR MIM; 607069; -;
 DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 KW Transmembrane.
 FT TRANSMEM 13 35 POTENTIAL.
 FT TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 98 120 POTENTIAL.
 FT TRANSMEM 127 149 POTENTIAL.
 FT TRANSMEM 164 183 POTENTIAL.
 FT TRANSMEM 190 208 POTENTIAL.
 FT TRANSMEM 223 245 POTENTIAL.
 SQ SEQUENCE 351 AA; 8CFBF66322DFEBFC CRC64;
 Query Match 1.0%; Score 9; DB 1; Length 351;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 385 ALLLLRLA 393
 DB 175 ALLLLRLA 183
 RESULT 3
 HIS4 AGRTS
 ID HIS4 AGRTS STANDARD; PRT; 247 AA.
 AC P58790;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase (EC 5.3.1.16)
 DE (Phosphoribosylformimino-5-aminomidazole carboxamide ribotide isomerase).
 DE HIS4 OR ATU0040 OR AGR_C_63.
 GN Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=11743193;
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Glendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the Biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J02817; AAA34273.1; -;
CC PIR: A29412; A29412.
CC HGSP; P00729; ICPY.
CC MEROPS; S10.009; -;
CC InterPro; IPR000379; Ser_estrs_site.
CC InterPro; IPR001563; Serine carboxpept. 1.
CC Pfam; PF00450; serine carboxpept; 1.
CC PRINTS; PR00724; CRBOXYPTASEC.
CC ProDom; PD001189; Serine carboxpept; 1.
CC PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
CC PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
KW SIGNAL 1 21
FT PROPEP 22 73 BY SIMILARITY.
FT CHAIN 74 484 SERINE CARBOXYPEPTIDASE III.
FT PROPEP 485 500 BY SIMILARITY.
FT ACT_SITE 216 216 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT BINDING 407 407 SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 500 AA; 55334 MW; B2ACB10BF8484CDA CRC64;

Query Match 0.9%; Score 8; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 ESVPEEP 410
Db 485 ESVPEEP 492
|||||
RESULT 7
ID NPPE_BACAM STANDARD; PRT; 521 AA.
AC P06832;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bacillolysin precursor (EC 3.4.24.28) (Neutral protease).
GN NPR.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RX MEDLINE=85006739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filpula D.;
RT "Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.";
RL J. Bacteriol. 159:811-819 (1984).
CC -1- FUNCTION: THERMOLABILE EXTRACELLULAR ZINC METALLOPROTEASE.
CC -1- CATALYTIC ACTIVITY: Similar, but not identical, to that of
CC thermolysin.
CC -1- COFACTOR: Binds 1 zinc ion and 4 calcium ions per subunit (By
CC similarity).
CC

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the Biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z95617; CAB09102.1; -;
CC EMBL: A007034; AAK45973.1; ALT_INIT.
CC PIR: D70985; D70985.
CC HGSP: P14779; IUFZ.
CC TIGR: MT1706; -;
CC TubercuList; RV1666C; -;
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 372 372 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47864 MW; DCBA4CF08FAD1A94 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 AADLSAL 637
Db 274 AADLSAL 281
|||||
RESULT 6
ID CBP3 WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735 (1987).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- INDUCTION: BY gibberellic acid (GA).

CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: K02497; AAB05346.1; -;
 CC PIR: A25415; HYBSN.
 CC HSP: P00800; ITRL.
 CC MEROPS: M04.014; -;
 CC InterPro: IPR005075; Pep M4 propep.
 CC InterPro: IPR001570; Peptidase M4.
 CC InterPro: IPR006025; Zn_MTPeptide.
 CC Pfam: PF03413; Pep_M4_propep; 1.
 CC Pfam: PF01447; Peptidase M4; 1.
 CC Pfam: PF02868; Peptidase M4 C; 1.
 CC PRINTS: PR00730; THERMOLYSIN.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Calcium; Zymogen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 221
 FT CHAIN 222 521
 FT METAL 364 364
 FT ACT SITE 365 365
 FT METAL 368 368
 FT METAL 388 388
 FT ACT SITE 449 449
 FT ACT SITE 521 521
 SQ SEQUENCE 521 AA; 56940 MW; 9609FA0416222751 CRC64;
 Query Local 0.9%; Score 8; DB 1; Length 521;
 Best Match Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 47 SVAAVASF 54
 Db 8 SVAAVASF 15
 |||||
 RESULT 8
 NPFE_BACSU STANDARD; PRT; 521 AA.
 AC P06142; P25268;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bacillolysin precursor (EC 3.4.24.28) (Neutral protease) (MCP 76).
 GN NPFE.
 OS Bacillus subtilis,
 OS Bacillus subtilis var. amylolacchariticus, and
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423, 1483, 1408;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC SPECIES=B.subtilis;
 RX MEDLINE=85006776; PubMed=6090407;
 RA Yang M.Y., Ferrari E., Henner D.J.;
 RT "Cloning of the neutral protease gene of Bacillus subtilis and the
 RT use of the cloned gene to create an in vitro-derived deletion
 RT mutation.";
 RT J. Bacteriol. 160:15-21(1984).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC SPECIES=B.subtilis;
 RX MEDLINE=97124187; PubMed=8969500;
 RA Winters P., Caldwell R., Enfield L., Ferrari E.;
 RT "The ampS-nprE (124 degrees-127 degrees) region of the Bacillus
 RT subtilis 168 chromosome: sequencing of a 27 kb segment and
 RT thermolysin.

RT identification of several genes in the area.";
 RL Microbiology 142:3033-3037(1996).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC SPECIES=B.subtilis; STRAIN=NS15-4;
 RA Lee S., Yoon K., Nam H., Chae K.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RC SPECIES=B.subtilis; STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [5]
 RN SEQUENCE FROM N.A.
 RC SPECIES=B.s.amylolacchariticus;
 RA Yoshimoto T., Oyama H., Takeshita T., Higashi H., Xu S., Tsuru D.;
 RT "Nucleotide sequence of the neutral protease gene from Bacillus
 RT subtilis var. amylolacchariticus.";
 RL J. Ferment. Bioeng. 70:370-375(1991).
 RN [6]
 RN SEQUENCE OF 1-158 FROM N.A.
 RC SPECIES=B.subtilis; STRAIN=168;
 RA Purnell B., Presecan E., Glaser P., Richou A., Danchin A.,
 RA Goffeau A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE OF 222-521.
 RC SPECIES=B.s.amylolacchariticus;
 RA Kobayashi R., Yoshimoto T., Tsuru D.;
 RT "Complete amino acid sequence of neutral protease from Bacillus
 RT subtilis var. amylolacchariticus.";
 RL Agric. Biol. Chem. 53:2737-2749(1989).
 RN [8]
 RN SEQUENCE OF 222-521.
 RC SPECIES=B.pumilus; STRAIN=76;
 RX MEDLINE=90148980; PubMed=2302386;
 RA Stoeva S., Kleinschmidt T., Mesrob B., Braunitzer G.;
 RT "Primary structure of a zinc protease from Bacillus mesentericus
 RT strain 76.";
 RL Biochemistry 29:527-534(1990).
 CC -!- FUNCTION: THERMOLABILE EXTRACELLULAR ZINC METALLOPROTEASE.
 CC -!- CATALYTIC ACTIVITY: Similar, but not identical, to that of
 CC thermolysin.

CC -!- COPACTOR: Binds 1 zinc ion and 4 calcium ions per subunit (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: K01985; AAC22627.1; -.
CC EMBL: AF012285; AAC24942.1; -.
CC EMBL: U30932; AAC2609.1; -.
CC EMBL: Z39111; CAB13343.1; -.
CC EMBL: D10773; BAA01604.1; -.
CC EMBL: Z97025; CAB09705.1; -.
CC PIR: A25414; HYBS.
CC PIR: JQ2129; JQ2129.
CC HSP: P00800; ITRL.
CC MEROPS: M04.014; -.
CC Subtilisin; B010448; nprE.
CC InterPro: IPR005075; Pep_M4_propep.
CC InterPro: IPR001570; Peptidase_M4.
CC InterPro: IPR006025; Zn_MTPeptidase.
CC Pfam: PF03413; Pep_M4_propep; 1.
CC Pfam: PF01447; Peptidase_M4; 1.
CC Pfam: PF02868; Peptidase_M4_C; 1.
CC PRINTS: PR00730; THERMOLYSIN.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Calcium; Zymogen; Signal;
CC Complete proteome.
CC
CC SIGNAL 1 27
CC PROPEP 28 221 POTENTIAL.
CC CHAIN 222 521 ACTIVATION PEPTIDE.
CC METAL 364 364 BACILLOLYSIN.
CC ACT SITE 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 368 368 BY SIMILARITY.
CC ACT SITE 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 449 449 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT SITE 449 449 PROTON DONOR (BY SIMILARITY).
CC CONFLICT 10 10 A -> R (IN REF. 1 AND 2).
CC CONFLICT 44 45 NA -> KP (IN REF. 1 AND 2).
CC CONFLICT 77 78 RL -> SV (IN REF. 1 AND 2).
CC SEQUENCE 521 AA; 56521 MW; 439E7B8F19D1E8F4 CRC64;
CC
CC Query Match 0.9%; Score 8; DB 1; Length 521;
CC Best Local Similarity 100.0%; Pred. No. 24;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 47 SVAAASF 54
CC DB 8 SVAAASF 15
CC
CC RESULT 9
CC ID_ZIC2 MOUSE STANDARD; PRT; 530 AA.
CC AC Q62520;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Zinc finger protein ZIC 2 (zinc finger protein of the cerebellum 2).
CC GN ZIC2.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OC NCBI_TaxID=10090;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=Cerebellum;
CC RC MEDLINE=96132843; PubMed=8557628;
CC

RA Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA Chapman V.M., Mikoshiba K.;
RA "The mouse zic gene family. Homologues of the Drosophila pair-rule
RA gene odd-paired";
RL J. Biol. Chem. 271:1043-1047(1996).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: CNS. A HIGH LEVEL EXPRESSION IS SEEN IN THE
CC CEREBELLUM.
CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D70848; BAA11115.1; -.
CC HSP: P08047; LSP2.
CC TRANSFAC: T04670; -.
CC MGD; MGI:106679; Zic2.
CC GO; GO:0007417; P:central nervous system development; IGI.
CC InterPro: IPR007087; Znf_C2H2.
CC Pfam: PF00096; zf-C2H2; 4.
CC ProDom: PD000003; Znf_C2H2; 1.
CC SMART: SM00355; Znf_C2H2; 4.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
CC Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein.
CC FT DOMAIN 20 23 POLY-HIS.
CC FT DOMAIN 25 33 POLY-ALA.
CC FT DOMAIN 89 97 POLY-ALA.
CC FT DOMAIN 227 231 POLY-HIS.
CC FT DOMAIN 232 239 C2H2-TYPE 1 (ATYPICAL).
CC FT ZN_FING 300 327 C2H2-TYPE 2.
CC FT ZN_FING 333 357 C2H2-TYPE 3.
CC FT ZN_FING 363 387 C2H2-TYPE 4.
CC FT ZN_FING 393 415 POLY-ALA.
CC FT DOMAIN 456 470 POLY-ALA.
CC FT DOMAIN 501 512 POLY-ALA.
CC SEQUENCE 530 AA; 55492 MW; 0065BD75B52E7DD2 CRC64;
CC
CC Query Match 0.9%; Score 8; DB 1; Length 530;
CC Best Local Similarity 100.0%; Pred. No. 24;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 478 SLAAQNG 485
CC DB 41 SLAAQNG 48
CC
CC RESULT 10
CC ID_ZIC2 HUMAN STANDARD; PRT; 532 AA.
CC AC Q95409; Q9H309;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Zinc finger protein ZIC 2 (zinc finger protein of the cerebellum 2).
CC GN ZIC2.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OC NCBI_TaxID=9606;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND VARIANT HPES POLY-ALA INSERTION.
CC RX MEDLINE=98442655; PubMed=9771712;
CC RA Brown S.A., Warburton D., Brown L.Y., Yu C.Y., Roeder E.R.,
CC Stengel-Rutkowski S., Hennekam R.C., Muenke M.;
CC "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila
CC

RT odd-paired.";
 RL Nat. Genet. 20:180-183 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20556339; PubMed=10984499;
 RA Yang Y., Hwang C.K., Junn E., Lee G., Mouradian M.M.;
 RT "ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine
 RL J. Biol. Chem. 275:38863-38869 (2000).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DISEASE: Defects in ZIC2 are a cause of holoprosencephaly type 5
 CC (HPE5) [MIM:603703]. HPE5 is a structural anomaly of the brain.
 CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF104902; AAC96325.1; -
 CC EMBL; AF193855; AAG28409.1; -
 CC HSP; P08047; 1SP2.
 CC TRANSFAC; T04237; -
 CC Genew; HGNC:12873; ZIC2.
 CC MIM; 603073; -
 CC GO; GO:0007420; P:brain development; TAS.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; zf-C2H2; 4.
 CC ProDom; PD000003; Znf_C2H2; 1.
 CC SMART; SW00355; Znf_C2H2; 5.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 CC Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 CC Disease mutation; Holoprosencephaly.
 CC FT DOMAIN 20 23 POLY-HIS.
 CC FT DOMAIN 25 33 POLY-ALA.
 CC FT DOMAIN 89 97 POLY-ALA.
 CC FT DOMAIN 226 230 POLY-HIS.
 CC FT ZN_FING 300 327 C2H2-TYPE 1 (ATYPICAL).
 CC FT ZN_FING 333 357 C2H2-TYPE 2.
 CC FT ZN_FING 363 387 C2H2-TYPE 3.
 CC FT ZN_FING 393 415 C2H2-TYPE 4.
 CC FT DOMAIN 456 470 POLY-ALA.
 CC FT DOMAIN 490 508 POLY-GLY.
 CC FT VARIANT 470 470 A -> AAAAAAAAAA (in HPE5).
 CC
 CC CONFLICT 124 128 /FTID=VAR_008856.
 CC SEQUENCE 532 AA; 55006 MW; BA3E6455DAF97EAC CRC64;
 CC
 CC Query Match 0.9%; Score 8; DB 1; Length 532;
 CC Best Local Similarity 100.0%; Pred. No. 24;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 478 SLAAQNG 485
 CC Db 41 SLAAQNG 48
 CC
 CC RESULT 11
 CC SBCC_RHOCA STANDARD; PRT; 1238 AA.
 CC AC O68032;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Nuclelease sbcD subunit C.

GN SBCC.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OX NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / St. Louis;
 RX MEDLINE=97404404; PubMed=9256491;
 RA Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
 RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter
 RL capsulatus SB1003.";
 CC Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388 (1997).
 CC -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
 CC can inhibit DNA replication and are intermediates in certain DNA
 CC recombination reactions. The complex acts as a 3'->5' double
 CC strand exonuclease that can open hairpins. It also has a 5'
 CC single-strand endonuclease activity (By similarity).
 CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF010496; AAC16118.1; -
 CC PIR; T03465; T03465.
 CC InterPro; IPR003439; ABC_transporter.
 CC Hydrolase; Nuclease; Exonuclease; DNA replication;
 CC DNA recombination; ATP-binding; Coiled coil.
 CC FT NP_BIND 37 44 ATP (POTENTIAL).
 CC FT DOMAIN 395 438 COILED COIL (POTENTIAL).
 CC FT DOMAIN 466 487 COILED COIL (POTENTIAL).
 CC FT DOMAIN 521 600 COILED COIL (POTENTIAL).
 CC FT DOMAIN 724 770 COILED COIL (POTENTIAL).
 CC FT DOMAIN 901 943 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1019 1052 COILED COIL (POTENTIAL).
 CC FT SEQUENCE 1238 AA; 128046 MW; 2B6BC6C63A859AE1 CRC64;
 CC
 CC Query Match 0.9%; Score 8; DB 1; Length 1238;
 CC Best Local Similarity 100.0%; Pred. No. 50;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 870 RAQALAQE 877
 CC Db 473 RAQALAQE 480
 CC
 CC RESULT 12
 CC HMEB_MXGSL STANDARD; PRT; 60 AA.
 CC ID_HMEB_MXGSL STANDARD; PRT; 60 AA.
 CC AC P31536;
 CC DT 01-JUL-1993 (Rel. 26, Created)
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Homeobox protein engrailed-like B (EN-B) (Fragment).
 CC OS Myxine glutinosa (Atlantic hagfish).
 CC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
 CC Myxiniidae; Myxiniinae; Myxine.
 CC NCBI_TaxID=7769;
 CC [1]
 CC RN SEQUENCE FROM N.A.
 CC RP TISSUE=Muscle;
 CC RX MEDLINE=9109509; PubMed=1980115;
 CC RA Holland P.W.H., Williams N.A.;
 CC "Conservation of engrailed-like homeobox sequences during vertebrate
 CC evolution.";
 CC RL FEBS Lett. 277:250-252 (1990).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59121; CAA41844.1; ALT_TERM.
 CC HSSP: P02836; 3HDD.
 CC InterPro: IPR000747; Engrailed.
 CC InterPro: IPR001356; Homeobox.
 CC InterPro: IPR000047; HTH_lambrepreser.
 CC Pfam: PF000046; homeobox; 1.
 CC PRINTS: PR00024; HOMEBOX.
 CC PRINTS: PR00031; HTHRPRESSR.
 CC ProDom: PD000010; Homeobox; 1.
 CC SMART: SM00389; Hox; 1.
 CC PROSITE: PS00027; HOMEBOX 1; 1.
 CC PROSITE: PS00033; ENGRAILED; PARTIAL.
 CC PROSITE: PS00071; HOMEBOX 2; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT NON_TER 1 1 HOMEBOX.
 FT DNA_BIND <1 41 HOMEBOX.
 FT NON_TER 60 60
 FT SEQUENCE 60 AA; 7044 MW; 3CE49F2081241292 CRC64;
 SQ
 Query Match 0.8%; Score 7; DB 1; Length 60;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qv 872 QALAEQ 878
 Db 23 QALAEQ 29
 RESULT 13
 ESAB MYCTU STANDARD; PRT; 94 AA.
 AC Q57155; O84901;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 6 kDa early secretory antigenic target (ESAT-6).
 GN ESAT6 OR RV3875 OR MT3989 OR MTV027.10.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1773, 1765;
 RX [1]
 RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection
 RT in mice."
 RL J. Immunol. 154:3359-3372(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=95204931; PubMed=7897219;
 RA Andersen P.; Andersen A.B.; Sorensen A.L.; Nagai S.;
 RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection
 RT in mice."
 RL J. Immunol. 154:3359-3372(1995).
 RN [2]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-10. AND CHARACTERIZATION.
 RC SPECIES=M.tuberculosis; STRAIN=Erdmann;
 RX MEDLINE=95247251; PubMed=7729876;
 RA Soerensen A.L.; Nagai S.; Houen G.; Anderson P.; Anderson A.B.;
 RT "Purification and characterization of a low-molecular-mass T-cell
 RT antigen secreted by Mycobacterium tuberculosis."
 RL Infect. Immun. 63:1710-1717(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Bahlmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basmah D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Emolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis;
 RA Singh B., Siddiqui Z., Singh S., Sharma P.;
 RT "ESAT-6 gene of a clinical isolate of Mycobacterium tuberculosis from
 RT India."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis;
 RX MEDLINE=96200095; PubMed=8631702;
 RA Mahairas G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;
 RT "Molecular analysis of genetic differences between Mycobacterium bovis
 RT BCG and virulent M. bovis."
 RL J. Bacteriol. 178:1274-1282(1996).
 RN [7]
 RP SEQUENCE OF 1-70 FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=99061212; PubMed=9846755;
 RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
 RA Gicquel B.;
 RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
 RT low-molecular-mass culture filtrate protein (CFP-10).";
 RL Microbiology 144:3195-3203(1998).
 CC -1- FUNCTION. NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY
 CC EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE
 CC RESPONSE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U34848; AAC4033.1; -;
 CC EMBL: X79562; CAA56099.1; -;
 CC EMBL: AL022120; CAA17967.1; -;
 CC EMBL: AF420491; AAL16896.1; -;
 CC EMBL: AE007190; AAK48357.1; -;
 CC EMBL: AF004671; AAC83446.1; -;
 CC FIR; A70803; A70803.
 CC TIGR; MT3989; -;
 CC Tuberculin; Rv3875; -;
 CC Antigen; Complete proteome.
 KW Tuberculin; Rv3875;
 FT INIT MET 0
 FT SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;
 SQ

```
Query Match          0.8%; Score 7; DB 1; Length 94;
Best Local Similarity 100.0%; Pred.No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 EAAASAI 339
DB 11 EAAASAI 17

RESULT 14
RLAI DICDI
ID RLAI DICDI STANDARD; PRT; 113 AA.
AC P22684;
DT 01-AUG-1991 (Rel. 19, Last Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S acidic ribosomal protein p1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AX3;
RX MEDLINE=91232920; PubMed=2030950;
RA Prieto J., Candel E., Coloma A.;
RT "Nucleotide sequence of a cDNA encoding ribosomal acidic
RT phosphoprotein p1 from Dictyostelium discoideum: identification of a
RT novel carboxy-terminal sequence in 'A' proteins."
RL Nucleic Acids Res. 19:1340-1340(1991).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X56193; CRA39656.1; -
DR PIR; S14013; R6DOP1.
DR DictyDb; DD01030; -
DR InterPro; IPR001813; 60s_ribosomal.
DR Pfam; PF00428; 60s_ribosomal; 1.
KW Ribosomal protein.
SQ SEQUENCE 113 AA; 11704 MW; D800B32325BAC301 CRC64;

Query Match          0.8%; Score 7; DB 1; Length 113;
Best Local Similarity 100.0%; Pred.No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 DDGIEIT 423
DB 20 DDGIEIT 26

RESULT 15
YJ91 ARCFU
ID YJ91 ARCFU STANDARD; PRT; 118 AA.
AC O28288;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1991.
GN AF1991.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
```

```

RN RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000965; AAB89266.1; -
DR PIR; F69498; F69498.
DR TIGR; AF1991; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 20 39 POTENTIAL.
FT TRANSMEM 46 63 POTENTIAL.
FT TRANSMEM 67 85 POTENTIAL.
SQ SEQUENCE 118 AA; 13011 MW; 824BD213B8A1FEBD CRC64;

Query Match          0.8%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred.No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 LIALLLL 389
DB 31 LIALLLL 37

Search completed: December 12, 2003, 17:39:42
Job time : 18 secs
```


QY 1 MPAGRLPRRCPPMMTKFTDCTRSNRIOPTTHRGYILKNNROIKLIASVAVAAASFOAHAGI. 60

QY 1 MPAGRLPRRCPPMMTKFTDCTRSNRIOPTTHRGYILKNNROIKLIASVAVAAASFOAHAGI. 60

Db 1 MPAGRLPRRCPCMTKFTDCTRSNRIQPPTHRGVILKNNRQIKLIAASVAVAAAFQAHAGL 60
 QY 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120
 Db 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120
 QY 121 VRDPVLVFRIGAGAVREYTAIILDPVGYSPKTKSALSDGKTHRTKTAFTAESQENONAKAL 180
 Db 121 VRDPVLVFRIGAGAVREYTAIILDPVGYSPKTKSALSDGKTHRTKTAFTAESQENONAKAL 180
 QY 181 RKTDDKDSANAANKVPAKNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 Db 181 RKTDDKDSANAANKVPAKNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 QY 241 HGRIRAGSVLHIPLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPKPEAKVAA 300
 Db 241 HGRIRAGSVLHIPLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPKPEAKVAA 300
 QY 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAASAIPTPTDETGNVSEPEVQVSAE 360
 Db 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAASAIPTPTDETGNVSEPEVQVSAE 360
 QY 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSQAARRTEESVPEEPDLDAA 420
 Db 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSQAARRTEESVPEEPDLDAA 420
 QY 421 EITAEVETPATPEAPKNDVNDTALDGESEELSAKQTFDVTETPSNRIDLD 480
 Db 421 EITAEVETPATPEAPKNDVNDTALDGESEELSAKQTFDVTETPSNRIDLD 480
 QY 481 AAQNGILSGALTQDEETQKADADWNAIESDVSVEPETNPNPVEIIVDTPEPESVAQ 540
 Db 481 AAQNGILSGALTQDEETQKADADWNAIESDVSVEPETNPNPVEIIVDTPEPESVAQ 540
 QY 541 TAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEQ 600
 Db 541 TAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEQ 600
 QY 601 TPEELHDFLKVYETDAVETAPETDFNAAADDLSALLQPAEAPSVENITETVAETPDF 660
 Db 601 TPEELHDFLKVYETDAVETAPETDFNAAADDLSALLQPAEAPSVENITETVAETPDF 660
 QY 661 NATADDLSALLQPAEAPSVENITETVAETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 720
 Db 661 NATADDLSALLQPAEAPSVENITETVAETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 720
 QY 721 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPAEAPSVENITETVAETPDF 780
 Db 721 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPAEAPSVENITETVAETPDF 780
 QY 781 PDSNTSEADALPDFLKGEEETVQWSIYLSSENPNNADTSPFSESVDGSDAPSEAKYDLA 840
 Db 781 PDSNTSEADALPDFLKGEEETVQWSIYLSSENPNNADTSPFSESVDGSDAPSEAKYDLA 840
 QY 841 BMYLEIGDRDAAAEVQKLEEAEGDVLKRAQALAEELGI 880
 Db 841 BMYLEIGDRDAAAEVQKLEEAEGDVLKRAQALAEELGI 880

RESULT 2

O86394 PRELIMINARY; PRT; 875 AA.
 AC O86394;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE TspA protein.
 GN TSPA.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=B:15:P1.16;
 RX MEDLINE=99307215; PubMed=10377136;
 RA Kizil G., Todd I., Atta M., Borriello S.P., Ait-Tahar K.,
 RA Ala'Aldeen D.A.A.;
 RT "Identification and characterisation of TspA, a major CD4+ T-cell and
 RT B-cell stimulating Neisseria-specific antigen.";
 RL Infect. Immun. 67:3533-3541(1999).
 DR EMBL: AJ010113; CAA09002.2; -;
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF01476; LysM; 1.
 SQ SEQUENCE 875 AA; 1F920E217A677091 CRC64;

Query Match 42.5%; Score 374; DB 2; Length 875;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 674; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MPAGRLPRRCPCMTKFTDCTRSNRIQPPTHRGVILKNNRQIKLIAASVAVAAAFQAHAGL 60

Db 1 MPAGRLPRRCPCMTKFTDCTRSNRIQPPTHRGVILKNNRQIKLIAASVAVAAAFQAHAGL 60

QY 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120

Db 61 GGLNIQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEA 120

QY 121 VRDPVLVFRIGAGAVREYTAIILDPVGYSPKTKSALSDGKTHRTKTAFTAESQENONAKAL 180

Db 121 VRDPVLVFRIGAGAVREYTAIILDPVGYSPKTKSALSDGKTHRTKTAFTAESQENONAKAL 180

QY 181 RKTDDKDSANAANKVPAKNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240

Db 181 RKTDDKDSANAANKVPAKNGKTHTVRKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240

QY 241 HGRIRAGSVLHIPLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPKPEAKVAA 300

Db 241 HGRIRAGSVLHIPLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPKPEAKVAA 300

QY 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAASAIPTPTDETGNVSEPEVQVSAE 360

Db 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAAASAIPTPTDETGNVSEPEVQVSAE 360

QY 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSQAARRTEESVPEEPDLDAA 416

Db 361 EETESGLFGGSYTLILLAGGGAALIALLLRLAQSQAARRTEESVPEEPDLDAA 420

QY 417 DDGIEITFAEVETPATPEAPKNDVNDTALDGESEELSAKQTFDVTETPSNRIDLD 476

Db 421 DDGIEITFAEVETPATPEAPKNDVNDTALDGESEELSAKQTFDVTETPSNRIDLD 480

QY 477 DSLAAQNGILSGALTQDEETQKADADWNAIESDVSVEPETNPNPVEIIVDTPEPE 536

Db 481 DSLAAQNGILSGALTQDEETQKADADWNAIESDVSVEPETNPNPVEIIVDTPEPE 540

QY 537 SVAQTAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKT 596

Db 541 SVAQTAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKT 600

QY 597 AEVQTPPELHDFLKVYETDAVETAPETDFNAAADDLSALLQPAEAPSVENITETVAE 656

Db 601 AEVQTPPELHDFLKVYETDAVETAPETDFNAAADDLSALLQPAEAPSVENITETVAE 660

QY 657 TPDFNATADDLSALLQPS 674

Db 661 TPDFNATADDLSALLQPS 678

RESULT 3

Q9K147 PRELIMINARY; PRT; 875 AA.
 ID Q9K147
 AC Q9K147;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

```

01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE TspA protein.
GN NMB0341.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciccione H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Knouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002391; AAF40784.1; -.
DR FICR; NMB0341; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01476; LysM; 1.
KW Complete proteome.
SQ SEQUENCE 875 AA; 92488 MW; 1F921520C167D090 CRC64;

Query Match 42.5%; Score 374; DB 16; Length 875;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 674; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MPAGLPRCPMMTKFTDCTRSNRIOPTTHRGYIILKNNRQIKLIAASVAASFOAHAGL 60
Db 1 MPAGLPRCPMMTKFTDCTRSNRIOPTTHRGYIILKNNRQIKLIAASVAASFOAHAGL 60
QY 61 GGLNQSLNDEPFSGSIITVGEAKALGGGVTVSEKGLTAKVHKLGDKAVIASSQA 120
Db 61 GGLNQSLNDEPFSGSIITVGEAKALGGGVTVSEKGLTAKVHKLGDKAVIASSQA 120
QY 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTSALSOGKTHRKTAAPTAEQENQAKAL 180
Db 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTSALSOGKTHRKTAAPTAEQENQAKAL 180
QY 181 RKTDKKDSANAARKPAYNGKTHTVKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTDKKDSANAARKPAYNGKTHTVKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRLRAGSVLHPNLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPEKPAKVA 300
Db 241 HGRLRAGSVLHPNLNRIKAEQPKPOTAKPAETASMPSEPSKQATVEKPEKPAKVA 300
QY 301 PEAKAEKPAVPEPVPAAATAAETAASAPQEAASAIPTDTGTNAVSEPEVQVSAE 360
Db 301 PEAKAEKPAVPEPVPAAATAAETAASAPQEAASAIPTDTGTNAVSEPEVQVSAE 360
QY 361 EETES-----GLFGSYTLILAGGGAALIALLLRLAOKKARTEESVPEEPDLDAA 416
Db 361 EETESGLFDGLFGGSYTLILAGGGAALIALLLRLAOKKARTEESVPEEPDLDAA 416
QY 417 DDGIEITFAEVETPATPEAPKPNVDNLALDGESEELSASQTFDVTDTPSNRIDLDF 476
Db 421 DDGIEITFAEVETPATPEAPKPNVDNLALDGESEELSASQTFDVTDTPSNRIDLDF 480
QY 477 DSLAAQNGILSGALTQDEETQKRAADWNAIESTDSVYEPETPNYPNVEIVDTPEPE 536
Db 481 DSLAAQNGILSGALTQDEETQKRAADWNAIESTDSVYEPETPNYPNVEIVDTPEPE 540
QY 537 SVAQTAENKPTVTDTSNDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKT 596
Db 541 SVAQTAENKPTVTDTSNDNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKT 600

RESULT 4
QY 597 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADLSALLQPAAPSVENITETVAE 656
Db 601 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADLSALLQPAAPSVENITETVAE 660
QY 657 TPDFNATADDLSALLQPS 674
Db 661 TPDFNATADDLSALLQPS 678

PRELIMINARY; PRT; 1065 AA.
Q8LJJO;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative leucine rich repeat containing protein kinase.
GN P0431H09.8
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0431H09.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003248; BAC10698.1; -.
DR Gramene; Q8LJJO; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007090; LRR plant.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR001245; Tyr kinase.
DR Pfam; PF00560; LRR; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00502; LRR_PS; 7.
DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 1065 AA; 115508 MW; 3E57C849FDA67DFF CRC64;

Query Match 1.1%; Score 10; DB 10; Length 1065;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 662 ATADDLSALL 671
Db 29 ATADDLSALL 38

PRELIMINARY; PRT; 141 AA.
Q9GRJ6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribosomal protein S12.
GN P214.46.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
```

RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RA Saunders D., Murphy L., Harris D., Ivens A.C., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL; AL449144; CAC14653.1;
 DR InterPro; IPR004038; Ribosomal_L7A.
 DR InterPro; IPR000530; Ribosomal_S1e.
 DR Pfam; PF01248; Ribosomal_L7Ae; 1.
 DR PRINTS; PR00972; RIBSOMAL512E.
 SQ SEQUENCE 141 AA; 15608 MW; ECF77D19AF538C74 CRC64;

Query Match 1.0%; Score 9; DB 5; Length 141;
 Best Local Similarity 100.0%; Pred. No. 5.1; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0;

Qy 732 EVPAVERN 740
 Db 9 EVPAVERN 17

RESULT 6

Q9A3G9 PRELIMINARY; PRT; 288 AA.
 ID Q9A3G9
 AC Q9A3G9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein CG3235.
 GN CG3235.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005987; AAK25197.1;
 DR TIGR; CC3235;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 288 AA; 29922 MW; BF3C71116A1F5E02 CRC64;

Query Match 1.0%; Score 9; DB 16; Length 288;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 AALIALLLL 389
 Db 128 AALIALLLL 136

RESULT 7

Q99986 PRELIMINARY; PRT; 396 AA.
 ID Q99986
 AC Q99986;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE VRK1.
 GN VRK1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Liver;
 RC MEDLINE=98008921; PubMed=9344656;
 RA Nezu J., Oku A., Jones M.H., Shimane M.;
 RT "Identification of two novel human putative serine/threonine kinases,
 RT VRK1 and VR K2, with structural similarity to Vaccinia virus B1R
 RT kinase.";
 RL Genomics 45:327-331(1997).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC EMBL; AB000449; BAA19108.1; -.
 DR HSSP; Q06486; 1CKI.
 DR Genew; HGNC:12718; VRK1.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_Thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 396 AA; 45476 MW; 5640C624BF059949 CRC64;

Query Match 1.0%; Score 9; DB 4; Length 396;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 824 SESVGSAP 832
 Db 59 SESVGSAP 67

RESULT 8

O29870 PRELIMINARY; PRT; 452 AA.
 ID O29870
 AC O29870;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Acetyl-CoA DECARBOXYLASE/synthase, subunit delta (CDHD).
 GN AF0377.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 DR EMBL; AE001078; AAB90859.1; -.

DR TIGR; AF0377; -
 DR InterPro; IPR004486; CdhD.
 DR Pfam; PF03599; CdhD; 1.
 DR TIGRFAMs; TIGR00381; cDhD; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 452 AA; 49747 MW; 55193972C432232A CRC64;

Query Match
 Best Local Similarity 1.0%; Score 9; DB 17; Length 452;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 730 PAEVPAAVEE 738
 DB 95 PAEVPAAVEE 103
 |||||

RESULT 9
 P72877 PRELIMINARY; PRT; 756 AA.
 AC P72877;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Exopolysaccharide export protein.
 GN EPSB OR SLL0923.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu B., Nakamura Y.,
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Ref. 3:109-136(1996).
 DR EMBL; D90901; BAA16893.1; -
 DR InterPro; IPR005479; CPase_I_D2.
 DR InterPro; IPR003856; LPS_Wzz_MPA.
 DR Pfam; PF02706; wzz; 1.
 DR PROSITE; PS00867; CPASASE_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 756 AA; 83640 MW; A5999B89C3EF7E58 CRC64;

Query Match
 Best Local Similarity 1.0%; Score 9; DB 16; Length 756;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 LLAGGGAAL 383
 DB 464 LLAGGGAAL 472
 |||||

RESULT 10
 Q93W58 PRELIMINARY; PRT; 1066 AA.
 AC Q93W58;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE NBS-LRR resistance-like protein J71.
 GN J71.
 OS Phaseolus vulgaris (Kidney bean) (French bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 OX NCBI_TaxID=3885;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Ferrier-Canu E., Geffroy V., Creusot F., Macadre C.,
 RA Imbert-Bollere P., Sevignac M., Langin T.;
 RT "Characterization and possible evolution history of expressed NBS-NRR
 RT R-like genes from the two major gene pools of common bean.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF306505; AAK61321.1; -
 DR EMBL; AF306501; AAK61317.1; -
 DR InterPro; IPR000767; Disease_resist.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR002182; NB-ARC.
 DR Pfam; PF00560; LRR; 1.
 DR Pfam; PF00931; NB-ARC; 1.
 DR PRINTS; PR00364; DISEASERSIST.
 SQ SEQUENCE 1066 AA; 122058 MW; 5385D3F299AAECC3 CRC64;

Query Match
 Best Local Similarity 1.0%; Score 9; DB 10; Length 1066;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AGLGGINTQ 66
 DB 699 AGLGGINTQ 707
 |||||

RESULT 11
 Q9XWP3 PRELIMINARY; PRT; 87 AA.
 ID Q9XWP3;
 AC Q9XWP3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Y43F8C.1 protein.
 GN Y43F8C.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; AL032637; CAA21608.1; -
 DR WormPep; Y43F8C.1; CE21899.
 SQ SEQUENCE 87 AA; 8841 MW; A78B1195507B4241 CRC64;

Query Match
 Best Local Similarity 0.9%; Score 8; DB 5; Length 87;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 LIALLLLL 390
 DB 4 LIALLLLL 11
 |||||

RESULT 12
 Q9HZ38 PRELIMINARY; PRT; 99 AA.
 ID Q9HZ38;
 AC Q9HZ38;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein PA3202.
 GN PA3202.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.

```

OX NCBI_TaxID=287;
RN [1];
RP SEQUENCE FROM N.A. / PA01;
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004744; AAC06590.1; --
DR InterPro; IPR005545; YCII.
DR Pfam; PF03795; YCII; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 99 AA; 10609 MW; 3AD945F44D54A85C CRC64;

Query Match 0.9%; Score 8; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 476 FDSLAAQAQ 483
Db 65 FDSLAAQAQ 72
|||||

RESULT 13
Q97AW7 PRELIMINARY; PRT; 121 AA.
AC Q97AW7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein TV0692.
GN TV0692 OR TVG0699210.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Yamazaki K., Kawamoto T.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP000993; BAB59834.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 121 AA; 13421 MW; 572FB9A3CF20BBAE CRC64;

Query Match 0.9%; Score 8; DB 17; Length 121;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 380 GAALIALL 387
Db 110 GAALIALL 117
|||||

RESULT 14
Q9Y9G2 PRELIMINARY; PRT; 131 AA.
AC Q9Y9G2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 131AA long hypothetical vacuolar ATP synthase subunit.

```

```

GN APE2326.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamaya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Nomura N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KI.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AF000064; BAA81338.1; --
DR InterPro; IPR002379; ATPase_Csub.
DR Pfam; PF00137; ATP-synt C; 1.
DR PRINTS; PR00122; VACATPASE.
KW Complete proteome.
SQ SEQUENCE 131 AA; 13380 MW; 314EB73BC2D8E17B CRC64;

Query Match 0.9%; Score 8; DB 17; Length 131;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 383 LIALLLLL 390
Db 122 LIALLLLL 129
|||||

RESULT 15
Q8DA80 PRELIMINARY; PRT; 143 AA.
AC Q8DA80;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN VW12329.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016804; AA010704.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 15812 MW; BC1F9A7C6AA9B8EC CRC64;

Query Match 0.9%; Score 8; DB 16; Length 143;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 383 LIALLLLL 390
Db 8 LIALLLLL 15
|||||

Search completed: December 12, 2003, 17:40:41
Job time : 46 secs

```